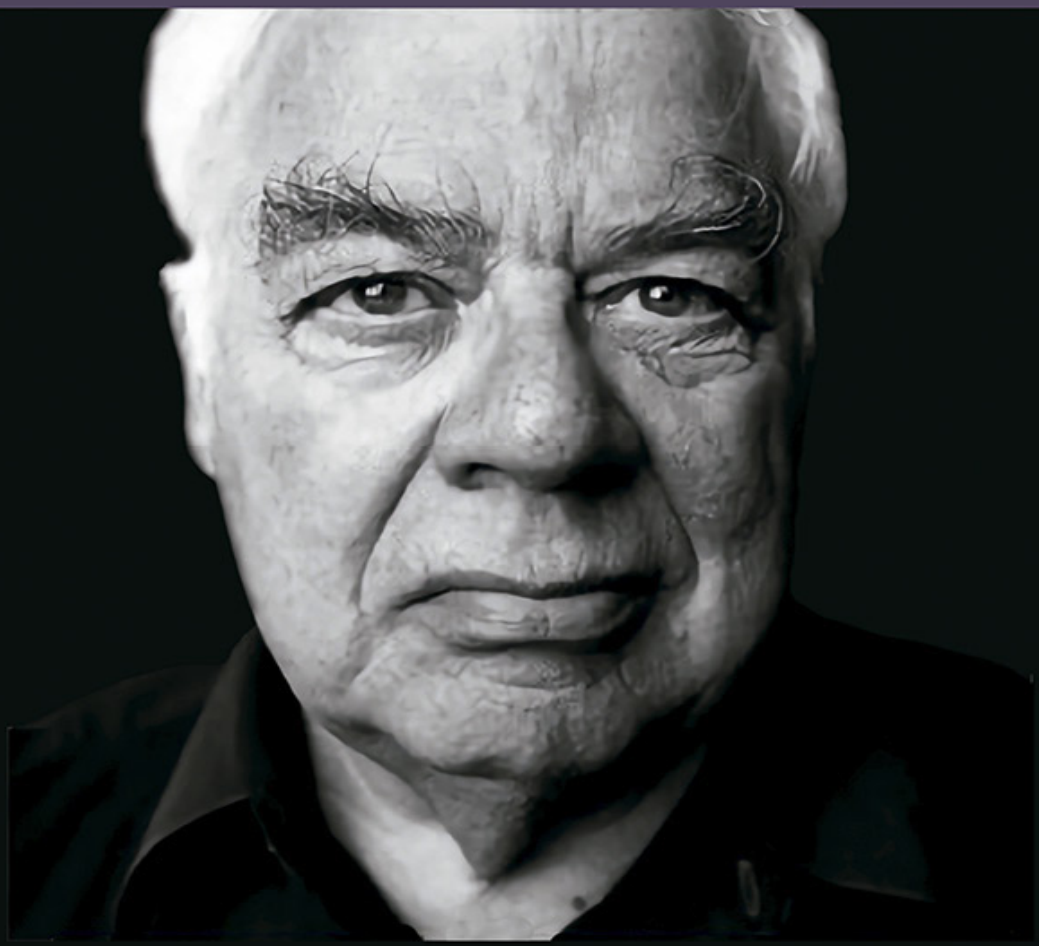
RICHARD RORTY

SOBRE LA FILOSOFÍA Y LOS FILÓSOFOS



Escritos póstumos

tecnos

RICHARD RORTY
W. P. MAŁECKI Y CHRIS VOPARIL
(Introducción y notas)

SOBRE LA FILOSOFÍA Y LOS FILÓSOFOS

ESCRITOS PÓSTUMOS

Traducción de ÁNGEL MANUEL FAERNA GARCÍA-BERMEJO

tecnos

ÍNDICE

AGRADECIMIENTOS

NOTA SOBRE LAS FUENTES

INTRODUCCIÓN. RORTY COMO FILÓSOFO CRÍTICO

I. ESCRITOS TEMPRANOS

Capítulo 1. La filosofía como ética

Capítulo 2. La filosofía como contemplación y participación

Capítulo 3. Kant como filósofo crítico

Capítulo 4. La paradoja del definicionismo

Capítulo 5. Reduccionismo

Capítulo 6. Fenomenología, análisis lingüístico y cartesianismo: comentarios sobre Ricoeur

Capítulo 7. La incomunicabilidad de las «cualidades sentidas»

Capítulo 8. Kripke sobre la identidad mente-cuerpo

II. ESCRITOS POSTERIORES

Capítulo 9. La filosofía como epistemología: respuesta a Hacking y Kim

Capítulo 10. La epistemología naturalizada y las normas: respuestas a Goldman y Fodor

Capítulo 11. La objetividad de los valores

Capítulo 12. Qué está muerto de Platón

Capítulo 13. El estado actual de la filosofía en Estados Unidos

Capítulo 14. El conversacionismo de Brandom: Davidson y *hacerlo explícito*

Capítulo 15. El hilemorfismo de McDowell y el naturalismo crudo

Capítulo 16. Reduccionismo versus semántica neo-wittgensteiniana

Capítulo 17. Observaciones sobre Nishida y Nishitani

CRÉDITOS

AGRADECIMIENTOS

Los editores queremos dar las gracias a varias personas que nos ayudaron a lo largo de los siete años de gestación de este volumen. Ante todo, a Mary Varney Rorty, que respaldó el proyecto desde el primer momento y lo hizo avanzar en cada etapa. En segundo lugar, al personal del Archivo de Teoría Crítica del departamento de Colecciones Especiales y Archivos de la Universidad de California en Irvine, donde se ubican los Richard Rorty Papers, especialmente a Steven McLeod y Audra Eagle Yun. También merece una palabra de gratitud nuestro colega y amigo Colin Koopman, que formaba parte del equipo editorial inicial que visitó por primera vez el archivo; debemos mucho a sus consejos y su agudeza crítica en estos años. Agradecemos el entusiasmo por el proyecto de nuestra editora en Cambridge University Press, Hilary Gaskin, y los útiles comentarios de los revisores anónimos de la propuesta. Richard Bernstein, amigo de Rorty toda su vida, nos aconsejó sabiamente y respondió preguntas con generosidad. Por último, nos gustaría reconocer la ayuda de Samuel Lelievre, Angel Lemke, Katarzyna Lisowska, Carlin Romano, Amie Thomasson y Jaroslaw Woźniak en la preparación del manuscrito para su publicación.

NOTA SOBRE LAS FUENTES

Los ensayos recogidos en este volumen se basan en documentos impresos y electrónicos de los Richard Rorty Papers, MS-C017, Special Collections and Archives, The UC Irvine Libraries, Irvine, California. Hasta donde hemos podido averiguar, Rorty no dejó instrucciones sobre el futuro de estos papeles, ni explicación alguna de por qué se quedaron en el cajón. A nuestro juicio, ese sino no se justifica por consideraciones de calidad: los textos cumplen de largo con el estándar general de los trabajos que Rorty dio a la imprenta en vida; algunos son tan buenos como lo mejor de lo que publicó. Creemos que el destino de la mayoría de estos textos ha tenido que ver con el hecho de que se escribieron para alguna ocasión especial, como congresos y conferencias. En esos casos, no habiendo planes de publicación inmediata, es probable que Rorty sencillamente pasara a tareas más urgentes y a nuevas invitaciones.

Con independencia de su grado de especialización, todos los ensayos requerían trabajo editorial en una u otra medida. Nuestras intervenciones no han sido nunca sustantivas. Normalmente, se han limitado a corregir errores tipográficos obvios y a añadir las referencias de las citas cuando faltaban. En tales casos, no lo indicamos en el texto.

Las mínimas intervenciones editoriales que hemos realizado al preparar los ensayos caen en dos categorías. Primero, y en consonancia con los volúmenes de escritos filosóficos de Rorty ya publicados por Cambridge University Press, hemos eliminado lo que son huellas obvias del contexto en el que se presentaba el trabajo. Así, en algunos manuscritos se refiere a diversos autores como el profesor X o Y («el profesor Ricoeur», por ejemplo), supuestamente porque esperaba que estuvieran presentes entre el auditorio. En otro manuscrito, se refiere al artículo de Alvin Goldman «The Relation Between Epistemology and Psychology» como «la ponencia de hoy»,

porque era parte del congreso en el que Rorty estaba exponiendo. «El profesor Ricoeur» se cambió por «Ricoeur», «la ponencia presentada hoy» por «The Relation Between Epistemology and Psychology», etc. En muchos manuscritos, usa contracciones propias del inglés hablado («I'm», «isn't», etc.), que por lo general no empleaba en letra impresa, y las hemos modificado en coherencia con su obra publicada. Estas intervenciones editoriales se han limitado a rasgos *obvios* de una presentación oral que podían cambiarse sin afectar al contenido y al tono. No hemos intervenido en el estilo o la sintaxis, que a veces son bastante informales.

El segundo tipo de intervención que requiere unas palabras de explicación se relaciona con la famosa costumbre de Rorty de dejar caer nombres, diciendo sin más que un determinado pensador (por ejemplo, Derrida, Foucault o Quine) sostiene *p* o *q* sin dar ninguna referencia a la correspondiente obra del autor o a su nombre de pila siquiera. Normalmente, sería tarea del editor localizar la fuente relevante y añadir el nombre propio del autor en cuestión cuando se cita por primera vez. Sin embargo, como la práctica de salpicar de nombres la redacción era un ingrediente característico del estilo de Rorty, casi siempre hemos optado por no hacerlo. Las excepciones han sido autores como, por ejemplo, Wolfgang Stegmüller, que, aunque seguramente le serían familiares al auditorio de ese momento, pueden resultar oscuros para una mayoría de lectores actuales. En tales casos decidimos añadir los nombres de pila y las referencias a las obras que más probablemente tendría en mente Rorty, usando «Eds.» para diferenciar nuestras notas al pie de las suyas. Allí donde los textos mecanografiados contienen correcciones o adiciones a mano de Rorty, las hemos tratado como si expresaran su intención en la versión final y las hemos incorporado al texto.

Algo debemos decir también sobre la datación de los textos. En algunos casos, el propio Rorty los fecha directamente en el manuscrito; en otros, la fecha se podía adivinar por la del evento en el que se presentó o por otras evidencias circunstanciales. A veces pudimos acotar un marco temporal, pero no determinarlo con precisión. Por ejemplo, no pudimos establecer exactamente cuándo escribió Rorty «Reduccionismo», pero, como ahí se centra en el libro

de J. O. Urmson de 1956, *Philosophical Analysis*, que trata como una obra reciente, y como alude a «lo que Bergmann llama el "giro lingüístico"», expresión que Bergmann usó por primera vez por escrito en una reseña de 1960, es probable que lo redactara a principios de los años 60 del pasado siglo. A continuación, damos la información que hemos logrado reunir sobre las fechas y las ocasiones de cada ensayo: en el caso de copias físicas, indicamos la caja [B, por «box»] y la carpeta [F, por «folder»] en que se localiza en los Richard Rorty Papers (por ejemplo, B44, F12); para archivos digitales, mencionamos su disponibilidad en la colección online «digital en origen» («b. d.») [por «born-digital»], que arranca en 1988. Varios textos en versión digitalizada están también en papel. Hay cinco ensayos que no llevan título en el original; los hemos titulado empleando expresiones que Rorty utiliza en ellos. Se trata de «La filosofía como ética», «La filosofía como contemplación y participación», «Fenomenología, análisis lingüístico y cartesianismo», «La filosofía como epistemología» y «El conversacionismo de Brandom».

1. «La filosofía como ética» (B44, F12) es un original a máquina sin fechar que hemos situado a principios de la década de 1960 o, quizá, a finales de la de 1950. La versión que se reproduce aquí incluye dos páginas escritas a mano por Rorty que aparecían unidas al texto mecanografiado.[1]
2. «La filosofía como contemplación y participación» (B44, F12) es un original sin título ni fecha, escrito a máquina en hojas sueltas y lleno de correcciones a mano. Como Rorty alude en él al «entusiasmo actual con el existencialismo», y sabemos que ya se interesaba por el existencialismo en la década de 1950, conjeturamos que muy probablemente el texto es de principios de los 60, o tal vez de finales de los 50.[2]
3. «Kant como filósofo crítico» (B44, F12; B8, F17), un original mecanografiado sin fecha y de la misma época que los dos textos anteriores, se compuso inicialmente como charla, con bastantes correcciones a mano del propio Rorty. En una

carpeta diferente, mecanografiada de nuevo para incorporar los añadidos a mano de la copia original, dimos con otra versión parcial, también sin fechar, de las seis primeras páginas.

4. «La paradoja del definicionismo» (B10, F14), un original a máquina parecido a los anteriores, data de finales de los 50 o primeros 60.[3] El manuscrito contiene números de llamada que envían a notas, pero por desgracia no hemos logrado localizar el texto de estas. Hemos conservado las llamadas que aparecen en citas y las hemos usado para dar la referencia de la cita; las demás se han eliminado.
5. «Reduccionismo» (B12, F1) es un original mecanografiado con correcciones menores a mano, probablemente de principios de los 60, como hemos explicado antes.[4]
6. «Fenomenología, análisis lingüístico y cartesianismo: comentarios sobre Ricoeur» (B13, F14) son las observaciones mecanografiadas de Rorty al trabajo de Paul Ricoeur «Husserl y Wittgenstein sobre el lenguaje», presentado en la reunión de la Sociedad de Fenomenología y Filosofía Existencial del 22 de octubre de 1966.[5]
7. «La incomunicabilidad de las "cualidades sentidas"» (B8, F8) se presentó en la reunión de 1968 de la División del Pacífico de la American Philosophical Association (APA).[6]
8. «Kripke sobre la identidad mente-cuerpo» (B8, F18) es el texto mimeografiado de una conferencia pronunciada en la State University de Nueva York en Binghamton en abril de 1973, con pequeñas correcciones a mano.[7]
9. «La filosofía como epistemología: respuesta a Hacking y Kim» (B12, F12) se preparó para un simposio sobre *La filosofía y el espejo de la naturaleza* de la División Oriental de la APA que tuvo lugar el 30 de diciembre de 1980. Wilfrid Sellars presidió el panel, e Ian Hacking y Jaegwon Kim presentaron ponencias, que posteriormente se publicaron.[8] La respuesta pormenorizada de Rorty no había aparecido impresa antes.

10. «La epistemología naturalizada y las normas: respuestas a Goldman y Fodor» (B13, F13) se presentó en una conferencia sobre «La psicología cognitiva y la nueva epistemología» celebrada el 6 de marzo de 1982 en la Universidad de Loyola de Chicago. La ponencia de Rorty incluye réplicas a las de Jerry Fodor y Alvin Goldman que la precedían en el programa.
11. «La objetividad de los valores» (b. d.) es un caso curioso, porque existe solo como archivo en versión digitalizada etiquetado con la fecha 25 de enero de 2000. No obstante, la evidencia textual interna sugiere una procedencia muy anterior; pensamos que data de 1981.[9]
12. «Qué está muerto de Platón» (b. d.; B12, F11) es una comunicación compuesta para un evento de febrero de 1988 con la participación del politólogo Dante Germino.
13. «El estado actual de la filosofía en Estados Unidos» (B7, F9) es una conferencia que Rorty fecha el 22 de septiembre de 1994. Las notas a mano al principio sugieren que fue pronunciada más de una vez.
14. «El conversacionismo de Brandom: Davidson y *Hacerlo explícito*» (b. d.; B13, F9) se compuso para una sesión «Encuentro del autor con sus críticos» sobre el libro de Brandom en la reunión de la División del Pacífico de la APA de 31 de marzo de 1995.
15. «El hilemorfismo de McDowell y el naturalismo crudo» (b. d.) se impartió en la Universidad de Columbia el 3 de agosto de 1995.[10]
16. «Reduccionismo versus semántica neo-wittgensteiniana» (b. d.) es un artículo compuesto en 2000.
17. «Observaciones sobre Nishida y Nishitani» (b. d.; B13, F21) es el texto, fechado el 1 de septiembre de 1999, de la conferencia introductoria pronunciada el mismo año en la Universidad Otani de Japón.

[1] Según las notas de archivo, los primeros tres trabajos de este volumen se encontraban originalmente en una carpeta de Rorty sin fecha y con la etiqueta «Conferencias populares». Véase *Guide to the Richard Rorty Papers MS.C.017:* www.oac.cdlib.org/findaid/ark:/13030/kt9p3038mq/dsc/, consultado el 19 de junio de 2019.

[2] Rorty señaló su interés por el existencialismo en un borrador de su solicitud para el programa de doctorado de Yale, al que accedió en 1952. Véase Neil Gross, *Richard Rorty: The Making of an American Philosopher*. Chicago, University of Chicago Press, 2008, p. 137 [hay trad. esp.: *Richard Rorty: la forja de un filósofo americano*. Valencia, Publicacions de la Universitat de València, 2010].

[3] En el trabajo, Rorty trata el debate de principios de los 50 sobre la llamada «paradoja de la predicción» como un tema vigente, califica de «reciente» un argumento que expone P. F. Strawson en un artículo de 1956, y no cita ninguna publicación posterior a 1956.

[4] Cf. el ensayo de Rorty de 1961 «The Limits of Reductionism», recientemente publicado en Richard Rorty, *Mind, Language, and Metaphilosophy: Early Philosophical Papers*, ed. de Stephen Leach y James Tartaglia. Nueva York, Cambridge University Press, 2014, pp. 39-54.

[5] Para el programa del simposio sobre «La filosofía del lenguaje: Husserl y Wittgenstein», véase *The Journal of Philosophy* 63, n.º 15 (1966), p. 452.

[6] Véase «News and Notes», *Philosophy East and West* 19, n.º 1 (1969), pp. 113-114. La lectura de Wittgenstein que desarrolla aquí Rorty está orientada a evaluar las perspectivas de Pitcher y Cook que discute en su artículo de 1970 «Wittgenstein, Privileged Access, and Incommunicability», reimpreso en *Mind, Language, and Metaphilosophy*.

[7] La correspondencia indica que Rorty fue invitado por Theodore Mischel (véase B8, F18).

[8] Véase Ian Hacking, «Is the End in Sight for Epistemology?», *The Journal of Philosophy* 77, n.º 10 (1980), pp. 579-588; y Jaegwon Kim, «Rorty on the Possibility of Philosophy», *The Journal of Philosophy* 77, n.º 10 (1980), pp. 588-597.

[9] Al final del texto, Rorty consigna como afiliación académica «Princeton University», donde enseñó desde 1961 hasta 1982, y se refiere al «difunto J. L. Mackie, de Oxford», que falleció en 1981.

[10] Véase Lance P. Hickey, *Hilary Putnam*. Londres, Continuum, 2009, p. 187.

INTRODUCCIÓN
RORTY COMO FILÓSOFO CRÍTICO

W. P. Małecki y Chris Voparil

Entre los filósofos influyentes, están los que publican poco y los que publican muchísimo. Si Ludwig Wittgenstein y Edmund Gettier pertenecen al primer grupo, en el segundo estarían, por ejemplo, Jacques Derrida, Jürgen Habermas y ese tábano de la filosofía norteamericana provocativo y prolífico que fue Richard Rorty. Rorty publicó a lo largo de su vida cientos de artículos y más de una docena de libros, tanto que cuesta trabajo imaginar que le diera tiempo a redactar algo de interés filosófico y que no conozcamos. Sin embargo, en el archivo Richard Rorty Papers de la Universidad de California en Irvine hay apiladas docenas de cajas con materiales de ese tipo. Algunas contienen notas para clases o primeros bocetos de ideas, pero en otras hay escritos filosóficos totalmente desarrollados, aptos para su publicación con un mínimo de edición. Este volumen presenta el material de esa segunda categoría.

Los escritos inéditos compilados aquí abarcan cuatro décadas de filosofía de Rorty. En los diecisiete ensayos que siguen, el lector hallará muestras de sus tempranas incursiones en la metafilosofía, a principios de los años 60 del pasado siglo, y de esa especie de *Geistesgeschichte* de largo aliento que él admiraba en Hegel y en Whitehead; parrafadas técnicas de seca argumentación típicas de su primera inmersión en la filosofía analítica del lenguaje y de la mente a mediados de los 60[1]; críticas de la epistemología y la metafísica tradicionales en su conocida vena anti-cartesiana y anti-kantiana; reflexiones sobre el estado de la filosofía en momentos históricos concretos; y, por supuesto, animados combates con sus amigos y enemigos filosóficos a lo largo de los años. Rorty se sentía sumamente cómodo en el formato del ensayo, y se nota. La prosa chispeante de aguda ironía, el ingenio vivaz, los hallazgos verbales, y

las interpretaciones a contracorriente que irritaron, instruyeron y deleitaron a generaciones enteras de lectores, aparecen aquí en todo su despliegue. Incluso cuando escribe sobre temas o pensadores de los que no se ocupó en su obra publicada, estas páginas ostentan el sello de la admirada y denostada voz autorial que hizo de Rorty el filósofo más leído y más glosado de su generación.

Estos textos destacan también por abrir ángulos nuevos, y a menudo más agudos, sobre cuestiones filosóficas que ocupan menos espacio o se tratan solo de pasada en su producción publicada. Entre los ensayos incluidos hay indagaciones sobre temas como el existencialismo, la incomunicabilidad de las cualidades sentidas, la objetividad de los valores o la epistemología naturalizada. Contienen también los que, hasta donde sabemos, son los únicos escritos que Rorty consagró en exclusiva a figuras como Platón, Kant, Sartre, Kripke y Ricoeur. Uno de 1999 presenta la que quizá sea su discusión más a fondo de filósofos no occidentales: Kitarō Nishida y Keiji Nishitani, de la Escuela de Kioto. El ensayo inicial, «La filosofía como ética», suministra un marco global que pone en continuidad preocupaciones que parecen emerger solo después en sus escritos publicados. Prácticamente todos los capítulos aportan iluminaciones importantes sobre la filosofía de Rorty; muchos aportan iluminaciones de interés filosófico general; y algunos hacen las dos cosas a la vez.

De entre los escritos que al mismo tiempo iluminan el propio pensamiento de Rorty y ofrecen contribuciones filosóficas genuinas, sobresale «Kant como filósofo crítico» (Capítulo 3). En este ensayo presenta un original concepto de filosofía crítica que le permite arrojar nueva luz sobre Kant, así como sobre la historia más amplia de la filosofía occidental. Lo logra describiendo interesantes analogías entre Kant, Aristóteles y Wittgenstein, que para él son por igual filósofos críticos. Dicho brevemente, el filósofo crítico trata de presentar un vocabulario que invalide, tanto los intentos escépticos de socavar un determinado dato como los intentos fundamentistas[2] de proporcionarle una base, donde el dato puede proceder de donde sea, desde «la conciencia moral común» hasta «la ciencia y las

matemáticas» de ese tiempo. El objetivo central de Aristóteles, Kant y Wittgenstein fue precisamente desarrollar un vocabulario de esta clase —un vocabulario dirigido a dos tipos de oponentes, los fundamentistas y los escépticos—, y esa es la razón de que sus filosofías provoquen reacciones parecidas. Resulta «mucho más fácil saber en contra de qué estaban, que saber de qué estaban a favor; es más —prosigue Rorty—, incluso las críticas a cada uno de sus respectivos oponentes son muy difíciles de formular. Cuando uno lo intenta, se ve diciendo cosas que no pueden conciliarse con las críticas al otro oponente respectivo».

Cualquiera que esté familiarizado con la recepción de la obra de Rorty reconocerá en esa clase de reacciones las réplicas que su propia filosofía suscitó durante décadas. Aquí la categoría de filósofo crítico nos permite entender mejor esa filosofía misma. Para empezar, aclara los comentarios positivos sobre Aristóteles y Kant que aparecen aquí y allá en su obra posterior, que de otro modo parecerían desconcertantes dada su conocida actitud negativa hacia los dos pensadores. Y, lo que es mejor, hace más nítido no solo el núcleo común que recorre toda su obra, sino también el importante cambio que se inicia a finales de la década de 1960 y culmina en *La filosofía y el espejo de la naturaleza*, dividiéndola en dos fases bien diferenciables.

Hubo un tiempo en que la literatura secundaria sobre Rorty describía ese cambio como el de un filósofo analítico ortodoxo que pierde la fe y se convierte de repente en pragmatista. Más tarde se vio que ese relato era inexacto. En realidad, Rorty se interesaba por el pragmatismo y estaba versado en él ya antes de hacerse filósofo analítico. Incluso en la época en que sus ensayos analíticos empezaron a obtener reconocimiento, su trabajo ya incorporaba premisas pragmatistas que solo después se volverían marca de la casa.[3] Desde entonces, los especialistas han venido discutiendo sobre la naturaleza exacta del cambio.[4] La categoría de filósofo crítico puede ser de ayuda en esos debates. O sea, se puede explicar la naturaleza del cambio que dio Rorty diciendo que, si bien era fundamentalmente un filósofo crítico tanto en su primera época

como en la segunda, en aquella perseguía una versión local de la filosofía crítica, mientras que en esta última se propuso una versión global.

Con «local» quiere aquí decirse que, en su etapa temprana, Rorty se embarcó en una filosofía crítica *dentro de* las fronteras de la filosofía tal como esta se había venido practicando tradicionalmente en Occidente. Propugnaba vocabularios que permitieran disolver ciertos intentos de socavar o fundamentar determinados datos, y veía eso como una contribución al avance de la filosofía. Este es el Rorty que escribía como un filósofo profesional y pensaba que la filosofía, tal como se practicaba en sus días, era importante. Lo que sucedió después fue que pasó a creer que la filosofía occidental tradicional no consistía nada más que en socavar y fundamentar, y que perseguir esas metas no merecía la pena en absoluto. Creía que lo más seguro era que fueran inalcanzables, ya que ningún filósofo había conseguido hasta el momento fundamentar o socavar nada, y que, incluso si pudieran alcanzarse, no ayudarían a resolver problemas sociales y políticos importantes.

Entonces, la filosofía crítica de Rorty se volvió global. En vez de ofrecer vocabularios pensados para disolver casos concretos de fundamentismo y escepticismo, empezó a ofrecer otros pensados para disolver el fundamentismo y el escepticismo en general, es decir, *globalmente*, y para disolver, por tanto, lo que él consideraba que era la filosofía tradicional como tal. Este es el Rorty que deja de argumentar como un filósofo profesional y que piensa que la filosofía, tal como se ha practicado tradicionalmente en Occidente, ha dejado de ser útil. Es el Rorty que aboga por una nueva filosofía con una meta y un método diferentes. La meta era coadyuvar a proyectos reales de fortalecer los vínculos democráticos, y ofrecer nuevas visiones del aspecto que tendrían esos proyectos comunitarios. El método sería el de la crítica cultural, una amalgama de antropología cultural, crítica literaria e historia. Este patrón de cambio desde una modalidad de filosofía crítica a la otra es visible en toda la obra publicada de Rorty, como lo es en los escritos compilados en este volumen.

I. ESCRITOS TEMPRANOS

En concreto, los dos primeros traslucen el tipo de intereses filosóficos que tenía Rorty antes de decidir ponerse a rueda de los debates analíticos en que tan absortos estaban sus nuevos colegas de Princeton.[5] Vemos ya la influencia del pragmatismo en su pensamiento. También es inconfundible su preocupación por la ineluctabilidad de elegir entre una pluralidad de puntos de vista alternativos. En «La filosofía como ética», tal vez la pieza más accesible del libro, remonta los orígenes históricos de la filosofía y el impulso inicial a filosofar al deseo de justificar valores. En un espíritu jamesiano, subraya la futilidad de esa «triste historia» de dos mil años tratando vanamente de justificar imperativos éticos. Con todo, Rorty halla aquí una lección positiva, al aducir que «una mala razón puede ser una buena historia». Utiliza la teoría de la verdad de James para mostrar que existe un modo pragmático de argumentar desde un «debe» a un «es» que esquiva los problemas asociados con el positivismo y el fundamentismo.

«La filosofía como contemplación y participación» (Capítulo 2), la única aproximación extensa de Rorty al existencialismo, revela la influencia que tuvo este movimiento sobre su pensamiento, que hasta ahora solo podía conjeturarse a partir de los temas y referencias existencialistas esparcidos por sus otros escritos.[6] Su tratamiento del distanciamiento y el compromiso como ideales en conflicto dentro de la filosofía no tiene solo un valor metafilosófico general, sino que resulta hoy especialmente oportuno, toda vez que a la filosofía se le reprocha, por una parte, el no contribuir lo suficiente a las luchas por la justicia social, y, por otra, el tener la osadía de pretenderlo siquiera. Rorty sostiene que «la tensión entre ser relevante y ser racional, entre inmediatez y mediación, entre participación y contemplación, es […] *el* problema filosófico», y subraya del existencialismo su «negativa a buscar seguridad a base de perseguir la objetividad» como una ruptura única en la historia de la filosofía.

«Kant como filósofo crítico» (Capítulo 3) es el único escrito de Rorty consagrado íntegramente a Kant, y su estudio más inquisitivo, original y empático sobre este pensador. Como hemos dicho antes, su novedoso enfoque sobre Kant descansa en un modo peculiar de entender la filosofía crítica, lo que le permite hacer tres cosas: primero, dar una explicación a la diversidad de interpretaciones de la obra de Kant; en segundo lugar, exponer sorprendentes analogías entre Kant y otras figuras de la historia de la filosofía a las que Rorty tipifica igualmente como filósofos críticos, como son Aristóteles y Wittgenstein; y, en tercer lugar, destacar el valor perdurable de las ideas de Kant, que para Rorty consiste en «hacer imposible» pensar que «la tarea de la filosofía es suministrar principios constitutivos que vayan a respaldar principios regulativos».

En «La paradoja del definicionismo» y en «Reduccionismo», Rorty empieza a aplicar sus ideas metafilosóficas de primera hora a cuestiones de filosofía lingüística prominentes en aquel momento.[7] En el primer texto (Capítulo 4) critica la tendencia dominante de la filosofía moderna a privilegiar, en metafísica y en epistemología, lo nítidamente definido sobre la indefinición borrosa, donde esta última se considera efecto de «la ignorancia y confusión» humana en lugar de verla, al modo de Aristóteles, como algo que existe en la naturaleza. Llama «definicionismo» a esta inclinación, y lo caracteriza como «la opinión de que no hay nada a lo que se pueda llamar razonablemente un enunciado y que no sea ni verdadero ni falso». «La resurrección del pragmatismo» —arguye Rorty—, con su contextualismo, «viene dándoles quebraderos de cabeza a los definicionistas». Pero también les invita a pasarse al bando pragmatista para evitar la paradoja a que se enfrentan, reconociendo la intuición de Dewey de que «toda transacción va a involucrar tanto elementos borrosos como no borrosos» y la idea de Peirce de la lógica «como una disciplina normativa y no descriptiva».

«Reduccionismo» (Capítulo 5) aborda la pregunta: «¿podemos abandonar el análisis reductivo como método de descubrimiento filosófico y, aun así, conservar los réditos intelectuales que ha devengado su uso como método para decidir qué cuestiones

discutir?» Tras presentar el programa de análisis lingüístico reductivo desarrollado en el siglo XX como una forma madura de la «concepción reduccionista» de la meta de la investigación que se tenía en el siglo XVII, Rorty analiza los argumentos de J. O. Urmson para concluir finalmente que este «no tiene en cuenta» aquellos «casos en que el análisis reductivo se aplica a los vocabularios técnicos de los filósofos». Aunque coincide con Urmson en que, juzgados por su propia vara de medir, la mayoría de los análisis reductivos fracasan, piensa que puede haber base para distinguir entre análisis útiles e inútiles. Vemos aquí también el temprano interés de Rorty en la eliminabilidad, que poco tiempo después se convertiría en el tema de una contribución específica.[8]

En «Fenomenología, análisis lingüístico y cartesianismo», Rorty aclara el porqué de su desinterés por la hermenéutica de Paul Ricoeur. Esta pieza se articula muy bien con el ensayo sobre Kant para presentar «la pugna entre fenomenología y análisis lingüístico como una competición entre dos grupos de oponentes a un mismo contrincante»: el cartesianismo. A su juicio, el mejor candidato de los dos para liderar la «revolución anti-cartesiana» es el analista lingüístico, cuya posición reduce a tres afirmaciones metafilosóficas centrales: la tesis Pragmatista, la tesis Naturalista y la tesis Convencionalista. Rorty se apoya en ellas para criticar la aproximación fenomenológica de Ricoeur a problemas de filosofía del lenguaje.

«La incomunicabilidad de las "cualidades sentidas"» (Capítulo 7) se centra en la tesis sobre la que han girado muchos debates generados por las *Investigaciones filosóficas* de Wittgenstein: la de que «no podemos comunicar ciertas cualidades» a otros —por ejemplo, la cualidad sentida específica del dolor de muelas. Rorty insinúa que los filósofos le han estado dando demasiada importancia a esa tesis. En cambio, él sostiene que solo es verdadera en un sentido «filosóficamente inocuo» —nunca podemos estar seguros de que, cuando describimos «X», lo que queremos decir o lo que sabemos sea lo mismo—, y que cuando se vuelve filosóficamente interesante es falsa, porque usar la expresión *dolor de muelas* de

manera correcta en circunstancias relevantes denota un conocimiento de la expresión incluso si nunca se hubieran experimentado las cualidades sentidas de un dolor de muelas. Al centrarse en el uso del lenguaje, amortigua la controversia filosófica y la amenaza del escepticismo epistemológico, para concluir que no necesitamos negar «ni la existencia de un sentido perfectamente válido de "saber" en el que puede haber conocimiento (o "percatación", o "consciencia") prelingüístico o no-lingüístico, ni la de particulares mentales no compartibles a los que tenemos acceso privilegiado».

En «Kripke sobre la identidad mente-cuerpo» (Capítulo 8), ofrece su estudio más extenso y sistemático de la obra de Saul Kripke. Pone los argumentos de este en un contexto histórico y filosófico que a menudo se pasa por alto y que arroja nueva luz sobre su viabilidad y su significado general. Todo aquel que esté interesado en la teoría de la identidad mente-cerebro, y en particular los que piensen que las críticas de Kripke demostraron que es insostenible y sigan perplejos por el hecho de que continúe viva y coleando, agradecerán las explicaciones de Rorty de por qué dichas críticas, en realidad, tienen que «dejar la cuestión sobre la identidad mente-cuerpo como estaba». Él predice que, tras las críticas de Kripke, «las viejas discusiones se traspasarán al nuevo vocabulario, hablando menos de significado y más de referencia, pero sin quebranto dialéctico para ninguna de las partes».

II. ESCRITOS POSTERIORES

En los ensayos que conforman la segunda parte del libro, los personajes, las concepciones y los argumentos que aparecen son similares, pero el tono resulta notablemente distinto, lo que refleja el paso de Rorty desde la versión local de la filosofía crítica a la global. Prácticamente todos hacen críticas a determinados proyectos dentro de la filosofía de su tiempo, pero las hacen principalmente para mostrar que la filosofía, según se practicaba en ese momento, tenía que abandonarse. Los proyectos son tanto fundamentistas como

escépticos, e incluyen los intentos de renovar la epistemología por parte de Hacking y Kim y de Goldman y Fodor (Capítulos 9 y 10), el escepticismo moral de J. L. Mackie (Capítulo 11), la modalidad de realismo de Crispin Wright y Bernard Williams (Capítulo 14), el platonismo naturalizado de John McDowell (Capítulo 15), la semántica reduccionista de «David Lewis, Saul Kripke, David Kaplan y John Perry» (Capítulo 16) y las modalidades de idealismo absoluto de Kitarō Nishida y Keiji Nishitani (Capítulo 17).

Rorty pretende mostrar que todos estos proyectos, por inteligentes e intrincados que sean, no hacen más que desperdiciar energía intelectual en algo carente de auténtico valor práctico. Y es así porque, insiste, los problemas que abordan son muy probablemente insolubles, e incluso, si se pudieran resolver, las soluciones no tendrían efecto alguno sobre la vida social. Pensemos en cuestiones como la de si el mundo puede «ejercer control racional sobre la mente», o la de si «los valores morales están "ahí afuera" o existen únicamente "en nosotros"» (Capítulos 11 y 14). Tales problemas solo serán importantes para quienes crean en doctrinas caducas como el hilemorfismo aristotélico o la «idea propia del siglo XVII de que aquello de lo que la ciencia no tiene conocimiento no es real», y lo mismo ocurrirá con sus soluciones. En este sentido, esos problemas y otros parecidos de la filosofía convencional son como los de la alquimia o los de la escolástica; habría que dejarlos atrás, y punto. En «La epistemología naturalizada y las normas», por ejemplo, Rorty admite que Goldman logra demostrar «la relevancia de la psicología para la epistemología», pero lo ve como algo parecido a haber probado con éxito «la relevancia de la física de partículas para la alquimia». Análogamente, en «El estado actual de la filosofía en Estados Unidos» (Capítulo 13) admite que muchos filósofos analíticos son «profesionales competentes, trabajadores, entusiastas», pero subraya que eso no significa de suyo que los problemas que abordan sean importantes. A fin de cuentas, entre los físicos del siglo XVII había muchas personas así discutiendo «cómo usar expresiones como "movimiento natural", "movimiento forzado", "cantidad de movimiento" e "inercia"»; o, entre los teólogos, discutiendo «cuántos

sacramentos instauró Cristo» (Capítulo 16). Sin embargo, los problemas en que se centraron parecen hoy simplemente estrafalarios.

En lugar de abordar esos problemas, Rorty querría que los filósofos contemporáneos contribuyeran al «fomento» de «los proyectos de cooperación social (por ejemplo, construir una sociedad igualitaria, sin clases, sin castas)» cultivando la crítica cultural (Capítulo 17). Trata de dar ejemplo involucrándose él mismo en la crítica cultural, en su caso «explicando a la gente por qué la filosofía no va a hacer algunas de las cosas que a veces se esperan de ella» (Capítulo 16). Las explicaciones que ofrece en la segunda parte del libro toman la forma de amplios relatos sobre la historia de la filosofía en general, y sobre la historia de la filosofía analítica en particular. Según él, la razón de que la filosofía analítica resulte socialmente inservible no es solo que se preocupa de un conjunto de problemas sin arreglo heredados de la tradición filosófica (véase «Qué está muerto de Platón»), sino también que se inventó como una forma de poner a la filosofía en «la senda segura de una ciencia» («El estado actual de la filosofía en Estados Unidos») y, de esa manera, separarla de la historia, la literatura y la política.

Aquí hay que subrayar, no obstante, que el cuadro que pinta Rorty de la filosofía analítica no es enteramente sombrío, pues considera que hay filósofos en ella que escapan a la mencionada dinámica. Se centra concretamente en Donald Davidson y Robert Brandom, situándolos en la tradición del pragmatismo filosófico y reconstruyendo la historia de este movimiento (véanse los Capítulos 13, 14, 15, 16 y 17). Ve su origen en el proyecto de los pragmatistas clásicos de combinar a Hegel y Darwin para desarrollar una «filosofía plenamente secularizada» que «contribuya a los esfuerzos por mejorar este mundo, en vez de posponer a otro la esperanza humana» («Observaciones sobre Nishida y Nishitani»). Y ve sus manifestaciones contemporáneas en los intentos de Davidson y Brandom de naturalizar la filosofía de la mente y la filosofía del lenguaje (Capítulos 14, 15 y 16). En «El conversacionismo de Brandom» (Capítulo 14), sostiene que esos intentos nos permiten romper con «la explicación platónica/aristotélica que distingue a los

seres humanos de los brutos por la capacidad de penetrar a través de las apariencias en la realidad», así como con toda una gama de debates inútiles que dicha explicación generó, como el de realistas contra antirrealistas o empiristas contra racionalistas. Por tanto, es en todo el arco de la tradición pragmatista donde busca Rorty las semillas del cambio que tanto necesita la filosofía.

Eso sí, uno puede considerar que ese cuadro se ha quedado obsoleto en muchos aspectos. Después de todo, ahora hay bastantes filósofos analíticos interesados directamente en problemas sociales urgentes. Pero se puede decir todavía que sigue siendo exacto en el sentido de que nos previene contra la creencia acrítica de que los filósofos pueden contribuir significativamente a solucionar problemas sociales de hoy trabajando sobre los problemas filosóficos de antaño, aquellos que nos legó precisamente la tradición filosófica tras la que va Rorty. Al tratar el problema social X o Y, para un filósofo siempre es tentador empezar por tratar la epistemología o la ontología de X o de Y, y acto seguido ponerse a dar vueltas sobre un conjunto de rompecabezas ontológicos y epistemológicos. Una lección que emana de los escritos de Rorty es que los filósofos deberían dejar de dar por sentado que sus indagaciones en la ontología o la epistemología de una cuestión social arrojará puntualmente resultados concretos, y que tales resultados, si se dan, pueden incidir lo más mínimo en cómo resolver en la práctica esa cuestión. Deberían ser más cautos que todo eso. Esta es solo una más de las importantes lecciones que, como esperamos haber mostrado, brindan los ensayos recogidos en este volumen.

---

[1] Para los trabajos tempranos de Rorty menos conocidos, véase *Mind, Language, and Metaphilosophy: Early Philosophical Papers*, ed. de Stephen Leach y James Tartaglia. Nueva York, Cambridge University Press, 2014.

[2] N. del T.: el inglés distingue entre «*foundationalist*» (el partidario de exigir fundamentos últimos para el conocimiento) y «*fundamentalist*» (el intransigente o fanático religioso). De ahí que sea más adecuado traducir el primer término como «fundamentista», para restringirlo al uso filosófico

relacionado con la idea de fundamento («*foundation*») y separarlo del político-religioso. En cuanto a «fundacionalismo», que también se emplea a veces, es obvio que confunde el diferente significado de «fundación» y «fundamento» en castellano.

[3] Véase, por ejemplo, Neil Gross, *Richard Rorty: The Making of an American Philosopher*. Chicago, University of Chicago Press, 2008 [hay trad. esp.: *Richard Rorty: la forja de un filósofo americano*. Valencia, Publicacions de la Universitat de València, 2010]. Para una explicación más completa que la que aquí podemos dar sobre la evolución filosófica de Rorty, véase *The Rorty Reader*, ed. de Christopher J. Voparil y Richard J. Bernstein. Malden (MA), Wiley-Blackwell, 2010.

[4] Es un cambio difícil de precisar bien porque no se trata de una ruptura abrupta, sino de una evolución en el tiempo. Ya en su introducción a *El giro lingüístico*, Rorty señalaba «dificultades» clave, inherentes al giro de la filosofía analítica hacia los métodos lingüísticos; dificultades que eran tanto metafilosóficas —esbozadas ya en sus ensayos de principios de los años 60— como epistemológicas, derivadas de la crítica por parte de «Dewey, Hampshire, Sartre, Heidegger y Wittgenstein» al modelo del espectador en teoría del conocimiento. Véase *The Linguistic Turn: Recent Essays in Philosophical Method*, ed. de Richard Rorty. Chicago, University of Chicago Press, 1967, p. 39 n. 75 y *passim* [hay trad. esp. de la introducción: «Dificultades metafilosóficas de la filosofía lingüística», en Richard Rorty, *El giro lingüístico*, Barcelona, Paidós / I.C.E.-U.A.B., 1990, pp. 47-133]. A medida que desarrollaba una explicación completa de esta crítica epistemológica y sus implicaciones, publicada finalmente como *La filosofía y el espejo de la naturaleza*, Rorty siguió expresándose a la manera analítica hasta mediados de los años 70 más o menos.

[5] El propio Rorty describió este esfuerzo de auto-educación profesional, subsiguiente a su llegada a Princeton en 1961, como «esmerarme en hacer de mí mismo una especie de filósofo analítico». «Si iba a ganarme el respeto de mis colegas —explicó— tenía que hablar de algunos de los temas que a ellos les preocupaban y escribir en una vena parecida a la suya». Véase Richard Rorty, «Intellectual Autobiography», en *The Philosophy of Richard Rorty: Library of Living Philosophers*, ed. De Randall E. Auxier y Lewis Edwin Hahn. Chicago, Open Court, 2010, pp. 13 y 11.

[6] Véase, por ejemplo, la discusión sobre Sartre y «la visión "existencialista" de la objetividad» en la tercera parte de *Philosophy and the Mirror of Nature*. Princeton (NJ), Princeton University Press, 1979 [hay trad. esp.: *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, 1983]

[7] Estas tesis aparecen más desarrolladas en su introducción a *El giro lingüístico*, momento en el que ya percibe las limitaciones del proyecto lingüístico.

[8] Véase «In Defense of Eliminative Materialism», *Review of Metaphysics* 24, 1 (1970), pp. 112-121.

I
ESCRITOS TEMPRANOS

CAPÍTULO 1
LA FILOSOFÍA COMO ÉTICA

Los filósofos padecen un extraño riesgo laboral: la gente se les acerca todo el rato y les pregunta por las buenas qué es lo que hacen y cómo lo hacen. No es algo que se pregunte a biólogos, economistas o músicos: de ellos la gente sabe bastante bien lo que hacen, y los detalles pueden interesarles o no. Pero el filósofo es diferente; cuesta mucho imaginar a qué dedica el tiempo.

En la primera mitad de esta exposición, quiero responder de manera muy aproximada a la pregunta de qué hacen los filósofos. Será, por fuerza, una respuesta histórica. Es así porque no existe tal cosa como «la filosofía» o «el filósofo»: hay sencillamente lo que diversas personas a las que se llamó filósofos dijeron o hicieron en diversas épocas. A medida que intento contestar a la pregunta, irán saliendo a la luz parte de los problemas que afrontan los filósofos a la hora de justificar su propia existencia. En la segunda mitad, trataré de indicar cómo creo yo que pueden evitarse algunos de ellos.

La filosofía en cuanto que distinta de la ciencia comenzó, en mi opinión, cuando los hombres[1] empezaron a tratar de justificar sus valores. A Tales, que se hizo inmortal por comentar en alguna ocasión que todo era agua, se le llama a veces el primer filósofo. Pero a mí me parece que el único honor que merece por ese comentario es el de haber fundado la ciencia. Es Sócrates, y no Tales, el primero en hacer preguntas filosóficas, porque fue él quien preguntó por primera vez y con toda seriedad cómo podía estar seguro de que los valores que abrazaba eran los correctos. Cuando Platón nos dijo que tenía que haber unas cosas llamadas Formas porque, de lo contrario, no habría esperanza de saber jamás si de verdad estábamos viviendo bien o no, dio la primera teoría distintivamente filosófica. Era filosófica porque era una respuesta a la pregunta de Sócrates, una pregunta por lo que debemos hacer.

No solo es ahí donde comienza la filosofía misma, sospecho que por ahí empieza también el interés de todo filósofo en ella. Quizá haya personas que se embarquen en el estudio de la filosofía porque les preocupa si el mundo externo existe realmente, o si pueden distinguir entre soñar y estar despiertos, o si está bien argumentar con silogismos, pero yo nunca he conocido a ninguna. Aunque, una vez puesto en marcha, se pueda quedar uno totalmente absorto en cualquiera de esos problemas, nunca se comienza por ellos. Comienzas por la ética; es decir, empiezas por darte cuenta de que no todo el mundo tiene los mismos valores que tú, y con la esperanza de encontrar la forma de justificar los tuyos. Ese descubrimiento puede producirse mientras te debates con algún dilema moral real, o al ver que a un amigo no le dice nada una novela que tú admiras, o de cualquiera de otras mil formas posibles. Eso que hay en nosotros y que hace que deseemos que cada parte de nuestra experiencia encaje con todas las demás, hace también que intentemos encajar nuestros valores, o nuestras metas, o nuestras actitudes hacia otras personas, en un esquema más amplio. Ese intento, el de justificar valores, metas y actitudes, es lo que creo que mueve a la gente a estudiar filosofía.

Pienso también que el resto de las denominadas ramas de la filosofía han salido de la ética. Por supuesto, en el desarrollo de la metafísica, la lógica y la teoría del conocimiento, la simple curiosidad y el asombro han jugado su papel; pero, aunque no puedo razonarlo aquí, yo diría que la forma en que esas materias surgieron de hecho en la historia de la filosofía muestra que el verdadero motivo que había detrás de ellas era la necesidad de justificar ciertos valores. No espero poder explicarlo aquí, pero lo puedo apoyar indirectamente haciendo notar que, en cuanto un conjunto de problemas deja de ser relevante para la ética, deja de verse como parte de la filosofía. Creo que eso puede percibirse en la historia del significado de la palabra «filosofía». A finales de la Edad Media, casi todo lo que hoy consideraríamos que son «las ciencias» —naturales y sociales— era parte de la filosofía. Poco a poco, y una por una, las que ahora llamamos ciencias fueron adquiriendo identidad propia y estableciéndose por su cuenta. A mi juicio, eso ocurría en el instante

en que se comprendía que el tema de la ciencia en cuestión no tenía relevancia directa para la elección entre valores en pugna. Por ejemplo, la física se convirtió en un campo de investigación independiente en el siglo XVII, no porque desarrollara un novísimo método o lograra una súbita claridad y precisión de las que la filosofía carecía, sino porque los hombres terminaron por ver, por ejemplo, que el problema de si la Tierra daba vueltas alrededor del Sol o viceversa era del todo irrelevante para el de cómo debían comportarse. La polémica sobre la teoría copernicana dejó de ser asunto de teólogos y filósofos y se convirtió en un asunto de los astrónomos cuando, y solo cuando, los teólogos y los filósofos comprendieron que la gloria de Dios se reflejaba igual de bien en la teoría copernicana que en la ptolemaica, y que adoptar un modelo del universo en que la Tierra se moviera en torno al Sol no tenía por qué cambiar realmente nuestros valores. Dicho de otro modo, dejó de ser una cuestión filosófica cuando dejó de tener implicaciones éticas.

La polémica decimonónica sobre la evolución es otro ejemplo de la misma situación. Mientras pareció que de «el Hombre desciende del mono» se podía inferir «el Hombre no tiene por qué no actuar como un mono», la evolución fue un problema filosófico, apto para que lo discutieran obispos y profesores de metafísica. Tras hacerse claro que semejante inferencia no era posible o no era necesaria, se convirtió en un problema de los zoólogos y la geología dejó de ser para siempre una parte de la filosofía.

Ese proceso de unas disciplinas separándose de la filosofía a medida que su irrelevancia para la ética se vuelve obvia lo hemos visto incluso en nuestro propio siglo. Lo que ahora llamamos psicología se enseñó en los departamentos de filosofía mientras la gente pensó que algún nuevo hallazgo sobre el funcionamiento del cerebro podría rebatir, o respaldar, lo que Platón había dicho sobre el conocimiento o lo que dijo San Agustín sobre la libertad. Cuando se hizo claro que los hombres se las verían exactamente con los mismos problemas de valores, en exactamente los mismos términos, no importa qué se descubriera en relación con su sistema nervioso, la psicología se mudó a los laboratorios, y los departamentos de

filosofía vieron recortados sus presupuestos para poder pagar esos laboratorios.

La última rama de la filosofía en declarar su independencia ha sido la lógica. La lógica todavía la enseñan filósofos, pero dentro de cincuenta años probablemente la enseñarán únicamente matemáticos. Mientras se creyó que la lógica trataba de las Leyes del Pensamiento, y que estas eran de alguna manera primas hermanas de las Leyes del Ser y de la Ley Moral, naturalmente que era una materia filosófica. Ahora que la lógica formal se ha convertido en gran medida en una cuestión de disponer símbolos de diferente manera y de inventar símbolos nuevos en términos de los cuales se puedan definir los viejos, las riendas las han tomado los matemáticos, que en eso son buenos. Lo interesante es que a los filósofos no les importa mucho. Una vez se hace claro que la lógica formal no es de mayor ayuda que la fisiología del cerebro para respaldar lo que queremos decir sobre lo correcto y lo incorrecto, los filósofos tienden a perder interés.

Prácticamente lo único que hoy puede reclamar el filósofo como materia suya propia es la ética misma, la metafísica y la teoría del conocimiento. (Dicho sea de paso, cuando hablo de «ética», uso el término como abreviatura de esa materia más amplia que incluye la estética, la filosofía política y las partes filosóficas de la teología moral, además de lo que solemos llamar ética.) Por desgracia, hasta la metafísica corre peligro, no porque nadie quiera arrebatársela a los filósofos, sino porque su hermana, la teoría del conocimiento, sigue intentando liquidarla. La teoría del conocimiento pregunta: ¿qué es la metafísica, si es algo diferente de las ciencias? Sabemos cómo obtiene el científico su conocimiento: lo hace por inducción a partir de hechos observables. Sabemos cómo obtiene el suyo el matemático: eligiendo unos axiomas y deduciendo teoremas de ellos. Pero la metafísica, ¿qué es lo que hace? No parece que use ni la inducción ni la deducción. Si queda todavía algún tipo de conocimiento acerca del mundo después de que todas las ciencias especiales hayan hecho suyas sus respectivas materias, dime solo cuál es y cómo propones que lo obtengamos.

Desde luego, en esta intervención no puedo entrar en todas las respuestas que se han dado a esa pregunta. Me limitaré a decir que ninguna le parece muy satisfactoria a la mayoría de los filósofos contemporáneos. El problema al que se enfrentan es: ¿qué queda exactamente de la filosofía si la metafísica es una disciplina tan dudosa como parece? La característica más destacada de los filósofos actuales es que por primera vez les preocupa seriamente esta cuestión. Ya no están seguros de tener una buena respuesta cuando la gente les pregunta qué hacen todo el día. En estas últimas décadas, entre los filósofos ha ido cundiendo más y más la sospecha de haber estado argumentando su propio paso al desempleo.

Creo que a los filósofos les gustaría replegarse a la ciudadela de la ética y contestar que ellos se ocupan de valores, mientras que todas las ciencias especiales se ocupan de hechos. Pero no pueden utilizar este eslogan sin exponerse a la pregunta: perfecto, así que os ocupáis de valores, pero ¿qué hacéis con ellos? Es de suponer que no los proclamáis sin más; se diría que los filósofos argumentan, no que se limitan a predicar. Por supuesto, tampoco los realizáis, como puede hacer el estadista o el artista. Así que, ¿qué hacéis con ellos?

Llegados a este punto, pienso que el filósofo desearía responder: «los justificamos; es decir, mostramos qué valores son los correctos y cuáles no, y argumentamos la diferencia». Pero, si dice eso, se topa con las mismas preguntas de teoría del conocimiento que liquidaron la metafísica. ¿Cómo propones realizar esa tarea? ¿Qué métodos especiales tienes para distinguir entre valores superiores e inferiores? ¿En qué están basados tus argumentos? Y aquí, una vez más, no nos ha llegado ninguna respuesta que resulte generalmente aceptable.

Ahora bien, hablando grosso modo, se pueden tener dos reacciones ante esta situación, la reacción que podríamos llamar «negativa» y la «positiva». Negativamente, uno puede decir que esta triste historia muestra simplemente que el emperador está desnudo, que la filosofía se revela como un intento absurdo de dar justificación a lo que no se puede justificar. La prueba evidente de que los imperativos éticos no se pueden justificar es el mero hecho

histórico de que, después de dos mil años intentándolo, lo único que parecemos haber descubierto es que prácticamente todo es irrelevante para su justificación. Si seguimos sin poder responder a las preguntas de Sócrates después de todo este tiempo, será que eso indica que estaba preguntando mal. Esta opinión sobre la naturaleza de la ética se resume en la definición de F. H. Bradley de la metafísica: «encontrar malas razones para lo que creemos por instinto».[2]

Esta visión de la ética se hace precisa y explícita en el célebre dicho de Kant de que «el "debe" no se puede inferir del "es"». En otras palabras, un imperativo moral no puede respaldarse o justificarse con ningún hecho o teoría acerca de hechos, científica, metafísica o religiosa, ni con nada salvo otro imperativo moral.

Por ejemplo, un argumento de la forma

(1) «Esta es la mejor alternativa, y por tanto debe ser la que desees» es válido porque en la frase «esta es la mejor alternativa» hay presente un «debe» implícito[3]. En realidad, la premisa y la conclusión son solo dos modos distintos de decir lo mismo. Pero un argumento como

(2) «Esto causará un dolor innecesario a todos los implicados; por tanto, no debes hacerlo» no es válido a menos que se añada la premisa extra «uno debe maximizar la felicidad», que expresa un imperativo. Análogamente, un argumento como

(3) «Dios quiere esto; por tanto, debes hacerlo» no es válido a menos que uno incluya una premisa que diga «debes hacer lo que Dios quiere».

Así pues, parece que a nuestro argumento histórico obvio sobre la imposibilidad de la ética le podemos agregar otro puramente lógico y del todo concluyente. Desde esta perspectiva, entonces, decir que la filosofía es el intento de justificar tus valores equivale a decir que la filosofía es imposible. Es la perspectiva que han popularizado los positivistas como A. J. Ayer, quien no ve otro papel para los filósofos

que el de usar la teoría del conocimiento como escoba para barrer todas las confusiones de los filósofos anteriores.

Sin embargo, también se puede adoptar una perspectiva positiva sobre esta historia de tentativas filosóficas de justificar elecciones éticas. Uno puede sostener que, mostrando que es baldío decir que se debe obedecer tal o cual imperativo moral debido a tales o cuales hechos, la ética se ha purificado. Puede pensar que la ética se ha liberado al fin de esfuerzos absurdos por mostrar que se deben aceptar determinados valores morales porque la Tierra es el centro del universo, o porque la ciencia no es capaz de entender el origen de la vida, o porque Estados Unidos tiene un nivel de vida elevado. Esta perspectiva positiva puede aceptar el dicho kantiano de que no es posible derivar imperativos de enunciados fácticos, pero, en vez de suponer que significa que nos podemos olvidar de la ética, subraya y reitera el hecho de que una elección ética es un acto de libre decisión, y que la virtud reside en la autenticidad de esa decisión. Saluda la buena nueva de que Dios ha muerto, tanto el Dios de la religión como la variopinta legión de sustitutos que le han inventado los filósofos desde Platón. Esta es, simplificando mucho, la perspectiva que adoptan existencialistas como Sartre, que no ve como una decepción la imposibilidad de justificar la ética, sino como una liberación, y que responde a la pregunta de Sócrates diciendo que nunca debería haberse formulado, igual que hacen los positivistas, pero con un tono de voz claramente distinto.

Hoy la filosofía se está dividiendo en dos bandos cuyas actitudes hacia la ética y hacia el futuro de la filosofía están representadas por esas dos reacciones al problema de la justificación de los imperativos éticos. Lo más difícil de encontrar en el panorama filosófico contemporáneo es un filósofo que intente dar una respuesta directa a la pregunta directa de Sócrates: «¿por qué debo adoptar estos valores como últimos en lugar de estos otros?»

Pero no es por falta de ganas de que tal filósofo exista. Desde todos lados se nos dice que ni los positivistas ni los existencialistas dan cuenta satisfactoriamente de la ética. Nos dicen que los positivistas se encogen de hombros demasiado alegremente ante la pregunta de Sócrates y no se toman en serio los problemas morales.

Los existencialistas, por su parte, parecen dar vía libre a la arbitrariedad absoluta y a la irracionalidad desbocada. Tenemos que emprender un «retorno a la razón», dice la gente, por el que la filosofía recobrará su autorrespeto, abrirá otra vez el viejo tenderete y, es de presumir, se decidirá por fin a decirnos si los valores que hemos elegido son correctos o equivocados, y por qué.

Ahora, mi propia posición en esta disputa es, a grandes rasgos, que coincido con las tres partes en lo que cada una opina sobre las demás. Coincido con los positivistas en que la irracionalidad consciente y deliberada de los existencialistas es peor que inútil, porque es potencialmente peligrosa. Coincido con los existencialistas en que los positivistas tienen la misma sensibilidad moral que un rebaño de vacas. Coincido con aquellos que llaman a un «retorno a la razón» en que justificar nuestras creencias morales sigue siendo algo que debemos intentar y que la filosofía es una materia que todavía merece la pena. Y sin embargo (y esto es lo que más me inquieta), coincido tanto con los positivistas como con los existencialistas en que el «debe» no puede derivarse del «es». O sea, estoy de acuerdo en que la elección entre imperativos morales alternativos sigue abierta no importa qué hechos se aduzcan o qué teorías metafísicas se adopten. Ante cualquier conclusión que pueda alcanzar cualquier filósofo sobre la naturaleza de la realidad, o sobre la naturaleza y poderes de la mente humana, la pregunta de qué debemos hacer seguirá abierta. Si alguien me demuestra de una vez por todas que soy un pecador a merced de un Dios iracundo, la cuestión de si debo rendirme a la voluntad divina o sufrir un heroico martirio por desafiarle sigue estando abierta. Si se me demuestra que soy una colección aleatoria de átomos, reunida por las indiferentes leyes de la física y condenada a disolverse sin dejar rastro cuando el sol estalle el martes que viene, sigo sin avanzar nada en la decisión de con qué valores vivir los cinco días que me quedan.

Esta tensión —la sensación de que la ética es a la vez vitalmente necesaria y absolutamente imposible— ha perseguido a la filosofía desde Kant, y desde entonces se ha vuelto más y más fuerte. Es obvio que no voy a resolverla en esta intervención, pero sí quiero

señalar lo que me parece el único camino que aún queda abierto. Paso ahora a la segunda parte de mi exposición, puede que más constructiva.

* * *

Miremos más de cerca el argumento de que el «es» no implica el «debe». Vamos a suponer que preguntamos: ¿vale también la conversa? Es decir, ¿tampoco podemos inferir nunca un enunciado de la forma «hay un tal y cual» a partir de uno de la forma «debe hacerse tal y cual»? A primera vista, se diría que es seguro que no podemos.

Por ejemplo, el argumento

(4) «Debemos amarnos los unos a los otros; por tanto, tiene que haber un Dios cuya voluntad sea que lo hagamos» es obviamente inválido. De igual modo lo es

(5) «Todos debemos dedicar la vida a la discusión dialéctica; por tanto, tiene que haber una Forma del Bien que podamos descubrir mediante esa búsqueda dialéctica».

Estos argumentos parecen ejemplos obvios de tomar el deseo por realidad.

Pero supongamos que, en lugar de inferir que «tiene que haber un tal y cual», lo que inferimos es que «debes creer que hay un tal y cual». Nótese que aquí no hay ningún error lógico. Hay un «debe» tanto en la premisa como en la conclusión; estamos infiriendo un imperativo a partir de otro imperativo. No obstante, lo habitual ha sido mirar con suspicacia estos argumentos. Considérese el siguiente:

(6) «Debe haber amor entre los hombres; por tanto, debes creer que hay un Dios que quiere que eso deba ocurrir».

Este tiene tan mala pinta como el que mencioné en primer lugar. Sin embargo, hay algunos argumentos con esa misma forma que

pintan un poco mejor. Pensemos en el siguiente:

(7) «Debes aceptar lo que nos dice la investigación científica; por tanto, debes creer que el hombre desciende del mono».

O el argumento paralelo que podría dar un fundamentalista[4] de los viejos tiempos:

(8) «Debes tener fe en las Escrituras, así que no debes creer que el hombre desciende del mono».

Ahora, ninguno de estos argumentos es una bobada o resulta lógicamente objetable, aunque puede que ninguno sea convincente. La razón de que no sean tonterías evidentes, como sí lo es el de que debemos creer en Dios porque debemos amarnos los unos a los otros, es que esta vez la premisa es un imperativo sobre las condiciones de la creencia.

Creer, después de todo, es un tipo particular de acción, y pueden existir —de hecho, existen— imperativos éticos sobre creencias, igual que sobre todos los demás modos de acción. Continuamente deducimos consecuencias, desde un imperativo general, sobre los tipos de creencias que debemos tener, o sobre los métodos que debemos seguir para elegir nuestras creencias, o sobre los valores en que debemos fijarnos cuando elegimos una creencia entre alternativas en pugna. No hay nada de malo en inferir desde ahí, con ayuda de premisas fácticas adicionales, un imperativo más particular que diga que debemos tener una cierta creencia particular.

Me parece que eso demuestra, entonces, que hay un género de imperativos éticos que sí permite que hagamos inferencias a enunciados sobre hechos. En otras palabras, creo que se puede argüir del «debe» al «es», que así lo hacemos y que debemos hacerlo. Porque yo sostendría que las expresiones «hay de hecho un tal y cual» y «debemos creer que hay un tal y cual» son sinónimas a todos los efectos relevantes aquí. Quizá esta identificación de los significados de ambas expresiones les choque por escandalosa. Si es así, permítanme que me defienda invocando la autoridad de William

James y la definición pragmática de la verdad, según la cual «esta oración es verdadera» no significa ni más ni menos que «esta oración expresa una proposición que debes aceptar». Como es obvio, el tiempo de mi intervención no me da para responder a las objeciones que pesan sobre la teoría de la verdad de James. Solo diré que, en mi opinión, todos los desarrollos que se han producido en filosofía desde la época de James a la nuestra sirven para respaldar, tanto las indagaciones de los positivistas sobre la naturaleza del método científico, como las ampliaciones que los existencialistas han hecho del eslogan de Kierkegaard de que «la verdad es subjetividad».

Si, por mor del argumento, suspenden ustedes momentáneamente su incredulidad en relación con la teoría pragmática, me gustaría explicar por qué creo que lo que vengo diciendo es relevante para el problema de justificar imperativos éticos. Afirmé que la razón de ser de la ética es justificar nuestra acción; o, más exactamente, justificar la elección de los imperativos por los que regularemos nuestras acciones. El concepto habitual de justificación es deductivo. Es decir, los filósofos siempre han intentado encontrar un enunciado con «es» del que poder deducir un imperativo ético. Ese enunciado suele ser una teoría metafísica, o científica, o religiosa, sobre la naturaleza de las cosas o del hombre. Como el «debe» no puede salir del «es», siempre han fracasado. Pero si han desistido de toda justificación salvo la deductiva, es porque la única alternativa parecía ser la inductiva, y es obvio que esa no va a servir. Pues la cuestión de si alguien obedece un imperativo, o incluso de si quiere obedecerlo, es irrelevante para la verdad del imperativo. Un imperativo que nos llame a «desterrar todo pensamiento en uno mismo» no queda refutado por el hecho de que nadie lo haya logrado jamás, ni quedaría refutado incluso si se demostrara que nadie lo ha intentado nunca siquiera. Los valores pueden seguir existiendo haya o no alguien que los siga o que repare en ellos.

Pero, si la ética no puede operar ni deductiva ni inductivamente, ¿qué puede hacer? Me parece que el que podamos inferir el «es» del «debe», aunque no a la inversa, nos da la respuesta a esa pregunta.

Podemos emplear para justificar imperativos éticos el mismo método que utilizó Milton para justificar el comportamiento de Dios hacia el hombre. Podemos contar una historia; la historia que se tiene que contar si nuestro imperativo es el correcto. Juzgado como un argumento, deductivo o inductivo, *El paraíso perdido* es una birria. Juzgado como historia, no está nada mal. Tomado como la historia que debemos creer si aceptamos los valores de Milton, es la mejor justificación de esos valores que podemos encontrar.

Pero aquí pueden ustedes decir que Milton al fin y al cabo solo hablaba simbólicamente, y que lo que a nosotros nos interesa son hechos literales. Sin embargo, yo digo que un análisis cuidadoso de los métodos de la ciencia muestra que también usamos imperativos éticos como premisas necesarias en el ámbito de los hechos científicos. Considérese la siguiente situación:

A: La Tierra gira alrededor del sol.

B: ¿Por qué debo yo creer eso?

A: ¿Por qué? Porque es un hecho, simplemente.

B: De eso nada, es una teoría que explica los hechos.

A: Eres un tiquismiquis, una teoría que explica los hechos es un hecho.

B: No, no soy ningún tiquismiquis. Hay un número casi infinito de teorías diferentes que los hechos involucrados aquí —diversos fenómenos ópticos, observaciones telescópicas, etcétera— respaldarán y que pueden explicarlos.

A: Pero tienes que admitir que, de esas teorías, el modelo copernicano del universo, en el que la Tierra gira en torno al sol, es con diferencia la más simple y la más elegante, y la que mejor armoniza con la física newtoniana.

B: Sí, lo admito, pero ¿por qué está tan bien la simplicidad, y la elegancia, y el armonizar con otras teorías? ¿Son estos valores últimos?

A: Bueno, desde luego son mis valores últimos cuando se trata de decidir qué voy a creer.

B: Pero no los míos. La imagen del mundo que resulta de considerar la Tierra —el hogar del hombre— como el centro del universo es la que satisface valores que para mí son mucho más preciosos que esos.

Ahora, lo que importa en este diálogo, la cuestión de con qué hablante está uno de acuerdo, es que arranca como una discusión sobre hechos literales y termina en una discusión sobre valores últimos. Todavía más importante: esta última no se puede resolver «volviendo a los hechos». Los hechos siempre dejan margen a la interpretación, y es un enunciado de deber lo que determina cuál de las interpretaciones posibles va a escoger uno. Este es un ejemplo del tipo de inferencia del «debe» al «es» de que vengo hablando. Tomas como premisa un imperativo ético, y de él deduces los hechos; es decir, deduces qué historia sobre los hechos debes creer. En todos los casos hay una infinidad de historias posibles, de las que debemos elegir una. Cuál elegimos está determinado por los propósitos que queremos que cumpla la historia, y nuestros propósitos están regulados por el imperativo ético bajo el que estamos operando.

En el sentido en que estoy usando el término «historia», la ciencia moderna en su totalidad es entonces una historia; una historia que satisface los valores de simplicidad, elegancia y capacidad de producir consecuencias tecnológicas prácticas. Si hablo por mí, yo acepto esa historia, porque creo que esos valores son últimos. Pero si se me pidiera que los justificara, todo lo que podría hacer es señalar el fruto de aceptarlos. Dicho de otra manera, señalaría de nuevo a la imagen científica del mundo. Este modo de justificar no es ni puede ser nunca concluyente. De hecho,

lógicamente es circular. Pero pienso que la lección de la historia de la filosofía es que es el único modo que tenemos. Cuando te piden que justifiques tus valores, lo único que puedes hacer es señalar el conjunto de teorías que tus valores te obligan a sostener.[5] Este es el principal punto que quería dejar sentado en mi exposición. Permítanme terminar extrayendo dos corolarios.

En mi opinión, la historia de la filosofía muestra una larga serie de inferencias válidas del «debe» al «es» disfrazadas de inferencias del «es» al «debe». Se me ocurre, por ejemplo, que lo que de verdad dice *La república* de Platón no es: «hay Formas; por tanto, debemos vivir la Vida de la Razón». Lo que dice es: «debemos vivir la Vida de la Razón; por tanto, debemos creer en las Formas»; o, si prefieren una formulación más suave, pero pragmáticamente equivalente, «por tanto, debemos actuar como si hubiera Formas». Como de natural buscamos justificaciones deductivas, tendemos a pensar en la Teoría de las Formas como si proporcionara «un fundamento metafísico para la ética»; pero en cuanto hablamos de esa manera, somos reos de mala lógica. Hablar así es mala lógica, y, como lo es, resulta una mala táctica si lo que de verdad nos interesa es convertir a Trasímaco. Hace aparecer lo que es más seguro como dependiendo de lo menos seguro. Lo que a Platón le preocupaba en último término —lo que sospecho que en último término ha preocupado a todos los filósofos sistemáticos— es recomendar un modo de vida. El conjunto de cosas que uno dice sobre cómo es el mundo —tu ciencia, o tu metafísica, o las dos juntas— es parte de un modo de vida. Pero esa parte no es el fundamento de todo él.

Por consiguiente, lo que realmente ha tenido lugar en la historia de la ética no es el paulatino reconocimiento de que es necesario prestar atención a los hechos. No es un proceso de ir distinguiendo hechos de valores poniéndolos en compartimentos separados con la etiqueta «debe» y «es», compartimentos separados por la imposibilidad lógica de ir de uno al otro. Eso es solo lo que parece la historia de la ética cuando se ve desde la perspectiva de un cierto conjunto de valores éticos: los del humanismo renacentista y el cientificismo de la Ilustración. Son valores que la mayoría de

nosotros compartimos, y por eso la imagen del mundo que alza un muro entre los hechos y los valores parece que fuera la única posible. Pero no todo el mundo los comparte; para hombres tan diferentes como Lenin y San Ignacio de Loyola, había valores considerablemente más importantes. En la visión que tenía San Ignacio de la Iglesia cristiana, y en la que tenía Lenin del proletariado emancipado, hallamos la justificación de valores que afirman el derecho a que se nos prohíba aceptar como creencias sobre el mundo y sobre el hombre los resultados de la ciencia. Hay escalas de valores en las que el «objetivismo» —ya se lo tilde de «secular» o de «burgués»— es un pecado. Las elecciones verdaderamente importantes sobre lo correcto y lo incorrecto son elecciones que imponen una de esas escalas y no otra.

---

[1] N. del T.: en los ensayos de esta primera época, Rorty emplea muy frecuentemente «hombre» y «hombres» con el valor genérico de «seres humanos»; ese uso desaparece en los de fecha más cercana, como ha ido ocurriendo en general en la expresión oral y escrita.

[2] F. H. Bradley, *Appearance and Reality: A Metaphysical Essay*, segunda edición. Nueva York, Macmillan, 1897, xiv. —Eds. [Hay trad. esp.: *Apariencia y realidad: ensayo metafísico*. Santiago de Chile, Ediciones de la Universidad de Chile, sin fecha.]

[3] N. del T.: hemos subsanado el lapsus del manuscrito, que aquí dice «explícito», no corregido en la edición en inglés. Como explican los editores Małecki y Voparil en la «Nota sobre las fuentes», su edición de los textos ha seguido un principio de intervención mínima, corrigiendo únicamente lo que consideraban errores tipográficos obvios. Eso ha hecho que casi nunca se hayan rectificado lo que parecen descuidos de Rorty igualmente obvios, como el de este caso. En la traducción hemos considerado que lo primordial era restituir el sentido cuando no cabían dudas al respecto, y así procederemos en lo sucesivo sin necesidad de indicarlo en nota. En todas las rectificaciones hemos consultado previamente con los editores del libro, tanto para saber si el problema procedía del manuscrito o de la transcripción, como para ratificar lo correcto de nuestra interpretación.

[4] N. del T.: véase la nota de traducción de la página 18.

[5] Tomado como observación psicológica, el comentario de Bradley acertaba al decir que la metafísica consistía en encontrar malas razones para lo que creemos por instinto, y creo que ese comentario puede extenderse al resto de la filosofía. Pero una mala razón puede ser una buena historia. Todo el espectro de la cultura —incluidas tanto la ciencia como la ética, y tanto la poesía como la metafísica— es para mí una serie de historias, buenas y malas, que diferentes gentes han suscrito en virtud de sus diferentes valores. Algunas de esas historias desempeñan otras funciones; la ciencia, por ejemplo, nos permite dominar el entorno. Pero desde un punto de vista lógico, en cuanto a su dependencia última de una evaluación implícita, están todas en pie de igualdad.

CAPÍTULO 2

LA FILOSOFÍA COMO CONTEMPLACIÓN Y PARTICIPACIÓN

En esta exposición quiero hacer dos cosas. En la primera parte dejaré algunos comentarios genéricos sobre la contemplación y la participación en filosofía; en la segunda, daré una visión rápida, pedante y contemplativa del existencialismo.

Últimamente hay una sensación en el aire de que la filosofía es demasiado importante como para dejársela a los filósofos, y de que necesita que la rescaten de ellos. Es parecida a la de que la literatura es demasiado importante para dejarla a los críticos literarios, la guerra, demasiado importante para dejarla a los generales, y la religión, demasiado importante para dejársela a los clérigos. Resulta difícil no simpatizar con ella. Sin duda existen ciertas elecciones de las que sería inhumano declinar toda responsabilidad pasándoselas a especialistas. A esa clase pertenecen, parece claro, las elecciones morales, estéticas, intelectuales y espirituales que acarrean las decisiones sobre la guerra, la literatura, la religión y la filosofía. El entusiasmo actual con el existencialismo se debe en parte a la impresión de que está escrito por y para los seres humanos, igual que otras filosofías anteriores, y no por y para especialistas, y de que, por tanto, ha rescatado de algún modo a la filosofía de los filósofos. Hay un sentimiento, en buena medida probablemente inconsciente, pero aun así bastante fuerte, de que la filosofía desde Aristóteles hasta Hegel ha sido una conspiración más bien turbia para convertir materias que son, propiamente y sin discusión, de incumbencia universal inmediata, en el tema de una doctrina esotérica o de una competición esotérica entre doctrinas esotéricas. El espíritu fullero que Kierkegaard detectó entre los herederos de Hegel se ha proyectado hacia atrás sobre el pasado y hacia adelante sobre el

presente, y así ha llegado a parecer que la entera historia de la filosofía ha sido una especie de juego. A la luz de la interpretación marxiana de la historia, nos acostumbramos a mirar atrás y ver la historia de la controversia teológica como una pía patraña, el reflejo de luchas reales por el poder social y económico. De manera parecida, a la luz de la ironía de los existencialistas nos estamos acostumbrando ahora a ver la historia de la filosofía como una historia de victorias vacías y derrotas vacías en un juego que, o bien se asemeja a las disputas teológicas —un pálido reflejo de luchas reales en la vida real—, o, todavía peor, constituye un mero pasatiempo para intelectuales hastiados del mundo.[1]

Ese sentimiento de que los filósofos profesionales nos la han jugado de algún modo, y de que han convertido en entretenimiento privado lo que en puridad es preocupación esencial de todos, se puede expresar también en términos de la diferencia entre participar y ser espectador. Nos parece que el filósofo, como el poeta, o el guerrero, o el que da testimonio de una fe religiosa, debe arriesgar; que lo que lo hace distinto a otros hombres, si es que va a ser distinto, es su disposición a aventurarse. Cuando decimos que los filósofos profesionales han traicionado a la filosofía, nos referimos a que han adoptado una actitud distanciada, como de espectadores, hacia esa aventura. Es como si los críticos literarios hubieran empezado a creerse que solo su escritura crítica es «realmente» literatura, y que los poemas y novelas que critican no son, en cierto modo, literatura de verdad, sino solo una forma rudimentaria de esos informes suyos. En todos estos casos, el truco consiste en partir del contraste obvio y evidente de por sí entre participantes y espectadores, y luego, mediante una adecuada y sutil redefinición de términos, hacer que parezca que ser un espectador es en realidad la forma suprema y más perfecta de participación. Los griegos se enfadaron muchísimo con Sócrates cuando utilizó por primera vez esa artimaña. Él argüía que, absteniéndose de toda participación en los asuntos políticos de Atenas (y limitándose a ir por ahí metiendo las narices), en realidad, de alguna extraña manera, estaba participando mejor y más altamente que los que gobernaban la

ciudad. Como saben ustedes, una acalorada opinión pública terminó por ejecutarle. Pero, para cuando llegó Hegel un par de milenios después, la gente estaba lo bastante hecha a la trampa como para que oírle decir que escribir historias de la filosofía era en realidad la forma mejor y más alta de filosofar sonara perfectamente razonable.[2] En tiempos de Hegel, justo antes de que viniera Kierkegaard y destapara el juego, a los filósofos les iba estupendamente. Se las habían apañado para convencer al resto del mundo de que su papel especializado de espectadores universales y su práctica de ciertas técnicas misteriosas de investigación eran esenciales para el bienestar general. En la práctica, habían ocupado el puesto de los chamanes y los sacerdotes como personas a quienes se paga por estar pendientes de lo que vagamente se designaba como «las cosas elevadas», de modo que el resto de la sociedad no tuviera que molestarse.[3]

Pues bien, esta manera de ver el contraste entre el existencialismo y la filosofía tradicional en términos de una reacción contra la filosofía entendida como deporte de espectadores, y de retorno a la participación, me parece hasta aquí muy sensata. Pero mis comentarios sobre el carácter trapacero de los espectadores sientan una tesis que hay que contrapesar con una antítesis. Es totalmente cierto que los filósofos traicionan su obligación cuando se convierten en meros espectadores, pero lo es también que la traicionan al dejar de serlo. Lo que hace de la filosofía un negocio tan complicado es que el filósofo supuestamente tiene que ser a la vez un aventurero, que arriesga su mente y su alma, y un adusto y distante entrometido que insiste en la objetividad, el rigor y el análisis crítico. A los filósofos se los mira con recelo en cuanto se declaran firmemente a favor de algo, porque su trabajo es ser «juiciosos», «imparciales» y «objetivos». Al fin y al cabo, son los espectadores de la eternidad y de todo tiempo, la gente que puede confiarse en que mantendrá la cabeza fría pase lo que pase. Incluso los que están encantados con el existencialismo —porque lo que dicen los existencialistas les suena mucho más «relevante para su vida» que lo que dicen los filósofos tradicionales— se ponen a veces

un poco nerviosos con él y empiezan a murmurar: «puede que sea emocionante, pero ¿realmente es filosofía?» En general, los filósofos están perdidos si lo que dicen es relevante para la vida, y si no lo es, están perdidos también. Si es relevante, se critica que sean demasiado «subjetivos», y si no lo es, se los condena por demasiado «desapegados».

Podemos ver, pues, que el contraste entre contemplación y participación, que aparece resaltado en el del existencialismo con la filosofía tradicional, no es cosa nueva; es solo la expresión actual de una tensión que siempre existió y de un dilema en el que todo filósofo se ha encontrado siempre. La tensión entre ser relevante y ser racional, entre inmediatez y mediación, entre participación y contemplación, es el problema más viejo de la filosofía. De hecho, es *el* problema filosófico. La razón última de que la filosofía sea algo en verdad aventurado es precisamente que intenta resolver el problema sin solución: el de ser al mismo tiempo intrépida y cauta, franca y astuta, estar totalmente convencida de que la verdad sobre determinado asunto es tal y cual, a la vez que atiende con genuino respeto a todos los que dicen que no lo es. Tratar de mantenerse a caballo entre esos opuestos supone un dilema constante en la vida de todo aquel que piensa sobre problemas filosóficos, y es el hecho de la vida con el que más familiarizado está el filósofo profesional. En la práctica, se traduce en que todo filósofo se halla atrapado en un conflicto de deberes: por un lado, el deber para con su propia humanidad de no permitirse ser distante, y, por el otro, su deber para con la sociedad de mantener la distancia. A menos que alguien mantenga la distancia, a menos que haya alguien que se quede en los márgenes y mire con sorna, el sentido histórico se perderá. Y allí donde el sentido histórico se pierde, la sociedad y la cultura se vuelven ciegas e inhumanas. El filósofo, sabedor de ello, se encuentra con que tiene que desempeñar dos papeles. Debe ser un filósofo en el mejor sentido de la palabra, el de un hombre que encara la experiencia humana cada vez desde cero y crea en el lenguaje una nueva expresión para ella. Pero a la vez debe ser un mero filósofo, un filósofo en el peor sentido de la palabra, un

pedante historiador de la filosofía, alguien con la fuerte sospecha de que no hay nada nuevo bajo el sol.

De este modo, el dicho de que «la filosofía es demasiado importante para dejársela a los filósofos» resuena en el alma de cada filósofo. Se corresponde con el sentimiento que tiene de que la vida es demasiado corta para desperdiciarla en un cínico hastío con el carácter repetitivo y aparentemente absurdo de la historia. Para él, ese dicho significa: «filosofar es demasiado importante para ser solo la historia de la filosofía». Justamente porque los filósofos profesionales admiten esto, es por lo que el existencialismo se está poniendo ahora de moda tan rápido también entre ellos, no solo entre los amateurs. Nadie sabe mejor que los propios filósofos lo acertado del verso de Shakespeare que cita Kierkegaard: «tenemos a nuestros filósofos para convertir en comunes y corrientes cosas sobrenaturales e inexplicables».[4] Nadie mejor que él sabe que la siguiente generación de filósofos descubrirá que cualquier cosa que haya dicho era un lugar común, y anticuado además, digno de un simple párrafo como mucho dentro de un manual de historia. Pero, junto con esta noción deprimente, el filósofo profesional tiene también la de que su trabajo es sacrificar sus propias visiones, su propia creatividad, sus propias vislumbres de cosas sobrenaturales e inexplicables, en el altar de ese mismo manual. Como filósofo y no como historiador de la filosofía —en el primero de sus papeles, no en el segundo—, el filósofo no es diferente de cualquiera y es ridículo hablar de él como un profesional. Encarar la experiencia humana y expresarla no es una especialidad profesional, es sencillamente lo que, mejor o peor, hace un ser humano, cada ser humano. El profesional, en cuanto que participante de la filosofía, tiene simplemente las responsabilidades humanas universales. Pero en su segundo papel, como historiador de la filosofía, como espectador de ella, tiene la responsabilidad especial de ser un buen espectador. Y esa responsabilidad entraña sacrificar su primer papel al segundo, sacrificarse a sí mismo a su función social especial. Tiene la responsabilidad especial de prestarse al mismo tipo de

tratamiento que él, en su condición de historiador de la filosofía, ha dispensado a otros.

Esto último viene a parar en lo siguiente: es totalmente cierto que la filosofía es demasiado importante para dejársela a los filósofos.[5] Pero lo que hay que dejarles a «los filósofos», en este sentido estrecho del término, es la posibilidad de actuar como espectadores, de hacer comunes y corrientes cosas que son sobrenaturales e inexplicables. Esa posibilidad hay que dejársela a ellos porque son los únicos que poseen las herramientas para hacerlo, es decir, los únicos que tienen la perspectiva histórica que te permite mantener la distancia y la cabeza fría pase lo que pase. La posibilidad lo es también —de hecho, es una invitación— para convertirse en un pedante y, a la postre, en un párrafo. Pero el ideal de un filósofo es el del hombre que acepta tal posibilidad con los ojos abiertos.

Permítanme que trate de reafirmar esto un poco más volviendo a la analogía entre el filósofo y el general. Cuando decimos que las guerras son demasiado importantes para dejárselas a los generales, tendemos a olvidar que los generales también son seres humanos, seres humanos llamados a sacrificar su propio sentido de la decencia humana cuando ordenan que los habitantes de una ciudad perezcan quemados bajo bombas incendiarias. Hacen ese sacrificio en nombre de su función pública, abandonan lo que tienen en común con otros hombres para hacer lo que creen que, a la larga, será un servicio a sus semejantes. Sin pretender decir que los filósofos profesionales afrontan el mismo tipo de decisiones morales desgarradoras que los generales (es del todo obvio que no), sí quiero insinuar que su sacrificio es análogo. El general sacrifica su sentido de la decencia humana; el filósofo, su sentido de un propósito, su sentido de seriedad. Los dos sacrifican algo de sí mismos en aras de una meta vagamente visualizada cuyo valor es antes social que individual. Ambos corren el riesgo de perder del todo su humanidad. La condición humana normal es la del participante, exprimir tus fuerzas en un empeño dotado de propósito, unificar tus acciones en un todo ético coherente. Es la posibilidad de esa cordura humana ordinaria lo que el filósofo sacrifica cuando adopta su segundo papel, su papel

como espectador de un drama histórico. Ese sacrificio, en la medida en que sea radical y completo, le impide tomarse en serio la filosofía y, por tanto, tomarse a sí mismo en serio. Lo empuja a una actitud más estética que ética hacia la filosofía, actitud que lo incluye a él mismo en su primer papel. Ahora bien, verse a uno mismo como objeto estético no es agradable; puede que sea lo más cerca que cabe estar de la locura conservando todavía la capacidad de comunicación articulada. Verte a ti mismo como un actor dentro de un drama inacabado, como un párrafo en potencia, es abandonar la introspección. Y, como dijo Kierkegaard, la ausencia de introspección es un tipo de demencia.

No obstante, después de pintar este desmesurado cuadro del profesional como hombre de gallardía y grandeza moral casi increíbles, debo repetir que lo que he descrito es un ideal. Probablemente no existen en la realidad generales ni filósofos que hagan semejantes sacrificios a plena conciencia. Para cuando llegue la situación en que pueda dar la orden de quemar vivos a los habitantes de una ciudad, el general ya se habrá convertido para protegerse en un técnico que no piensa y no ve nada más que el efecto de un bombardeo sobre producción enemiga. Se hará menos que humano a base de dejar de ser consciente de la necesidad de elegir. Sin embargo, si hace eso demasiado a menudo, puede que nunca sea capaz de refrenarse, que nunca vuelva a ser humano. Así ocurre también con el filósofo profesional, a una escala infinitamente menor, cuando ve que, para exponer adecuadamente su punto, debe excogitar unos cuantos párrafos aseados que aglomeren y resuman el trabajo de una docena de colegas sumamente importantes y sumamente heterogéneos. Lo hace sin pensar. Por lo general, y a dios gracias, se las arregla para olvidar qué está haciendo. Juega el juego por el juego mismo, confiando en que tenga un sentido, al mismo modo irreflexivo en que, para su propia protección, el general confía con los ojos cerrados en que la guerra tenga un sentido.

Una vez dichas las cosas genéricas que quería decir sobre los filósofos como participantes y como espectadores, intentaré volver a los negocios y hablar de existencialismo. Mi trabajo es tomar el existencialismo y reducirlo a unos cuantos párrafos de, como si

dijéramos, un manual de historia de la filosofía aún por escribir. En otras palabras, trataré de poner el existencialismo en algún tipo de contexto histórico, lo que inevitablemente significa domesticarlo y bajarle los humos. Todos mis comentarios hasta aquí no han sido otra cosa que una disculpa por intentar hacer eso. Quienes creen que el existencialismo ha salvado a la filosofía de los filósofos a menudo nos instan a que salvemos de ellos al existencialismo. Lo que se temen es que lo transformen por arte de magia en lo más aburrido que hay, en «solo una filosofía más». Dado que voy a hacer todo lo posible por convertirlo en solo una filosofía más, quería decir algo en defensa de este empeño. Mi defensa consiste básicamente en decir que, en el sentido en que el existencialismo no es un deporte de espectadores, ninguna otra filosofía lo ha sido tampoco. Pero, en otro sentido, y mirado desde el punto de vista del historiador de la filosofía, tiene tanto de espectáculo y, por ende, de objeto de especulación, como cualquier otra filosofía.

* * *

Lo más obvio que hay que decir del existencialismo es que tiene algo que ver con la subjetividad como opuesta a la objetividad. No es solo que los autores existencialistas hablen mucho y explícitamente de subjetividad, y de qué cosa estupenda es, sino que toman prestados los términos cruciales que utilizan de lo que solemos considerar nombres de sentimientos subjetivos. Mientras que los filósofos chapados a la antigua discutían problemas confortables, bien conocidos y de largo rodaje, sobre el espacio y el tiempo, la mente y la materia, la sustancia y la causalidad, el pensamiento y el ser, el libre albedrío, la existencia de Dios, y otros por el estilo, los filósofos existencialistas ignoran esas nobles cuestiones. En lugar de ello, dedican el tiempo a disquisiciones como la diferencia entre el temor y la angustia, la naturaleza de lo pastoso en tanto que distinta de lo viscoso, y la relación entre vergüenza y náusea. Los libros filosóficos antes llevaban títulos bonitos y directos como «Proceso y realidad», «La experiencia y la naturaleza», «Lenguaje, verdad y lógica», y demás. Pero ahora es frecuente que

tengan títulos desasosegantes, como «Muerte y libertad» o «Vergüenza y remordimiento». La costumbre del existencialismo de usar tales términos como categorías fundamentales ha llevado a algunos a tratar de desprestigiarlo negando que sea filosofía en absoluto y reduciéndolo a una simple modalidad de psicología introspectiva particularmente pretenciosa. A otros detractores los lleva a afirmar que es solo un intento camuflado de decir en mala prosa didáctica lo que únicamente puede decirse, si acaso, con la ficción o con la poesía. En mi opinión, ambos intentos de desacreditarlo van completamente descaminados. Tan ridículo es criticar al existencialismo por tomar sus categorías de la introspección psicológica y de la literatura, como criticar a Platón por tomar las suyas de las matemáticas, o a Aristóteles y Whitehead por extraerlas de la biología, o a Hobbes por hacer lo propio con la física. Es verdad que los filósofos se critican unos a otros por este tipo de cosas; Aristóteles, por ejemplo, acusó a Platón de no ser realmente un filósofo, sino solo un matemático particularmente presuntuoso. Pero no deberían hacerse ese tipo de críticas. La filosofía tradicionalmente ha consistido en tomar aspectos de la experiencia como cruciales, o como paradigmáticos, e intentar interpretar el resto de la experiencia a la luz de esos aspectos seleccionados. No es una crítica decirle a un filósofo que ha tomado prestadas sus categorías de alguna disciplina extra-filosófica; al fin y al cabo, ¿de dónde iba a sacarlas si no? La crítica solo puede ocuparse de si es buena su elección, y solo averiguamos si lo es o no examinando el producto acabado, es decir, mirando cómo se ve la totalidad de la experiencia a la luz de los aspectos seleccionados que el filósofo ha elegido como categoriales.

No obstante, parecería que, ya antes de examinar el producto acabado, al existencialista se le puede achacar que términos como angustia y vergüenza son elecciones obviamente malas. Porque parece obvio que, en filosofía, lo que queremos son categorías con un radio de aplicación lo más amplio posible. Eso es lo que hacía que categorías como espacio, causalidad o sustancia resultaran tan adecuadas. No hay prácticamente ningún ítem de experiencia del que no se pueda pensar que está en el tiempo, que tiene una causa,

y que es una sustancia o una propiedad de una sustancia. En cambio, hay montones de ítems de experiencia que no parecen tener nada que ver con fenómenos como la angustia, el vértigo o el estar arrojado. Además —podría proseguir el argumento—, uno debe elegir sus categorías de modo que se pueda distinguir tajantemente qué cosas caen en cada una. Un juego de categorías filosóficas debe ser como los cajones de un secreter bien diseñado, sin ambigüedades sobre dónde va cada cosa. Por ejemplo, si eliges el espacio como una de tus categorías, en seguida sabes que las mesas de escritorio son espaciales y las ideas no, no hace falta romperse la cabeza. Pero, con categorías tomadas de la introspección, no hay forma clara de distinguir qué entra en qué. Si usted y yo hablamos sobre la diferencia entre Dios y el Mundo, o entre la mente y la materia, creemos estar seguros de que los dos sabemos bastante bien de qué habla el otro cuando emplea esas palabras. Pero cuando hablamos sobre la diferencia entre la angustia y el temor no existe esa seguridad, y cuesta mucho ver cómo podríamos llegar a alcanzar alguna. Por consiguiente, cuesta mucho ver cómo podríamos ponernos de acuerdo nunca sobre qué se supone que entra en cada encabezado. Más en general, si lo que perseguimos en filosofía es lograr un modo de pensarlo todo que sea aceptable para todo el mundo, parece obvio que no debemos darnos a nosotros mismos la desventaja inicial de elegir categorías que significan cosas diferentes para diferentes personas.

He puesto en juego este argumento porque creo que oír la respuesta que le da el existencialismo es un modo eficaz de captar las percepciones básicas que subyacen a la revuelta existencialista. El argumento que acabo de delinear hace dos suposiciones cruciales. En primer lugar, supone que, a cuantos más ítems de experiencia quepa aplicar un término dado, más probabilidades tiene de ser candidato al estatus de categoría filosófica última. La segunda suposición es que, cuanto más acuerdo haya entre la gente sobre el significado de un término, más probabilidades tiene su candidatura. Las dos podrían parecer verdades autoevidentes. Pero lo parecen solo si uno tiene en mente un determinado modelo de lo que debe ser la filosofía. Lo parecerán si uno piensa que la filosofía imita el

procedimiento de las matemáticas, y que ofrece argumentos deductivos a partir de verdades incuestionables y prosigue hacia una infinidad de verdades más particulares, verdades particulares cuyo radio es tan amplio como el de los ítems de experiencia mismos. Si eso es explicar filosóficamente, si la filosofía es una especie de súper-ciencia en la que buscamos leyes universales que expliquen casos particulares, entonces ciertamente necesitaremos el tipo de categorías que se puedan aplicar en todas partes y resulten inteligibles a todos.

Ahora, es justo esta preconcepción de la naturaleza de la explicación filosófica lo que los existencialistas quieren desafiar. Con ello, desafían también los dos supuestos a que acabo de referirme sobre cómo elegir categorías. No obstante, antes de intentar enunciar el desafío que les lanza el existencialista, quiero remarcar cuán básicas son en la tradición filosófica esa preconcepción y los supuestos que le hacen de corolario, y dar pistas de lo profundamente incrustadas que están en el modo en que normalmente nos representamos la filosofía.

Para verlo, tenemos que retroceder a la pregunta de por qué filosofar en absoluto. Si el filosofar tradicional consiste, como he dicho, en aislar algunos aspectos de la experiencia y reducir otros aspectos a esos, ¿por qué queremos hacerlo? ¿Por qué es tan importante reducir lo Múltiple a lo Uno, o, como mínimo, los Demasiados a los Algunos Menos? Ahora bien, la respuesta obvia a esto parecería ser: es más fácil tratar con las cosas en esa forma reducida. Pero ¿qué quiere decir tratar con ellas? ¿Qué clase de tratos son esos? ¿Qué es exactamente lo que se vuelve más fácil si tenemos una explicación reductiva? ¿Qué sacamos de conocer las relaciones entre categorías universales como aquello que dicta las relaciones entre los hechos particulares?

Lo que se vuelve más fácil, está claro, es controlar la experiencia. Si podemos reducir la rebosante variedad de la experiencia a una serie de ejemplificaciones de categorías últimas, estamos en mejor disposición de saber cómo hacer que la experiencia sea lo que queremos que sea. La filosofía como súper-ciencia tiene la misma meta que la propia ciencia: el control sobre el entorno. El motivo

original que hay detrás del apremio a explicar reductivamente es la búsqueda de seguridad, de protección. Si podemos hacernos con un juego de categorías universales cuyas interrelaciones prefiguren el comportamiento de todos los hechos reales y posibles, entonces podemos estar a salvo de sorpresas; de hecho, podemos dejar de pensar y descansar al fin. Vista de esta manera, la historia entera de la filosofía tradicional, desde Platón hasta los pragmatistas, está unida por el acuerdo en un solo punto: a saber, que la meta de razonar es hacer innecesario el razonamiento y, por consiguiente, que la meta de filosofar es escapar a la situación en que filosofar es necesario.

Si miramos la filosofía a esta luz, podemos ver por qué los dos supuestos mencionados son comunes a todos los filósofos tradicionales. Hay dos peligros fundamentales a los que nos enfrentamos: el primero es que las cosas hagan algo inesperado y que no podamos controlar, y el segundo es que otro ser humano sepa algo que nosotros no sabemos. Ambos miedos contribuyen a nuestro sentimiento de inseguridad, y la filosofía, si va a hacer su trabajo, debe eliminarlos. Eliminará el primero si logra encontrar categorías que no sean meramente personales o locales, sino que lleguen tan lejos como el propio radio de la experiencia. Eliminará el segundo si logra formular las relaciones entre esas categorías en premisas que ningún ser humano pueda negar y que todos los seres humanos estén forzados a conceder. El resultado ideal de la investigación filosófica se alcanzará cuando no quede nadie que pueda oponernos un argumento, cuando tengamos un fundamento inamovible para todo lo que decimos, y cuando sobre cualquier cosa tengamos algo que decir. Por eso queremos categorías que todos puedan entender y cuyo significado sea claro para todos, porque solo así podremos asegurarnos de que todo el mundo percibe qué inamovible es el fundamento sobre el que descansamos; solo así podemos sentirnos seguros.

Pues bien, otra forma de llamar a ese tipo de seguridad es «objetividad». Decir que la meta de la filosofía es la seguridad y decir que es la objetividad son dos maneras de decir lo mismo. La palabra «objetividad» envuelve con sus dos sentidos las dos metas

gemelas del filosofar. Tener conocimiento objetivo, en cuanto que opuesto a subjetivo, es conocer el objeto tal como es, no como parece; es tomar el mando de las cosas y no estar a merced de sus acciones sobre nuestros sentidos y sobre nuestras demás facultades. Por otro lado, ser objetivo, en cuanto que opuesto a ser prejuicioso, o místico, o provinciano, es poder y querer dar razones de todo lo que uno dice, y hacer entender a todos los demás lo que uno piensa y por qué. Creemos que las ciencias son objetivas en la medida en que cumplen esas dos exigencias. Por eso creemos que las matemáticas son más objetivas que la física, la física, más que la biología, la biología, más que la economía, la economía, más que la sociología, y así sucesivamente. Cuando llegamos a la filosofía, creemos que es perfectamente objetiva, que tapa todos los agujeros de nuestro conocimiento y suministra a las ciencias subordinadas un fundamento metafísico o metodológico inamovible. La filosofía en cuanto que súper-ciencia se supone que establece principios generales de los que pueden derivarse por ejemplificación las leyes científicas, o bien, por lo menos, deja perfectamente claros los métodos, la estructura y el valor de las ciencias especiales, de forma que no queden cabos sueltos en el tejido de nuestro pensamiento. La búsqueda de una perfecta objetividad por parte de la filosofía debe convertir idealmente nuestro conocimiento en un sistema perfectamente cerrado, y así cerrarnos a nosotros perfectamente, darnos perfecta seguridad.

Este análisis del significado de la objetividad nos pone por fin en disposición de darle algún sentido al cliché de que el existencialismo hace hincapié en la subjetividad, no en la objetividad. Lo que viene a decir el cliché es que hace hincapié en la inseguridad en detrimento de la seguridad. Es bastante cierto, pero si eso fuera todo lo que hace el existencialismo, no sería más que otro intento —uno de muchos en la historia de la filosofía— de espolearnos a que nos aseguremos más a base de recordarnos lo horrible que es estar inseguros. Lo importante del existencialismo es que rompe con el concepto de la filosofía como un medio hacia la seguridad. Le da la espalda del todo a la idea misma de la filosofía como una súper-ciencia. Simplemente no ve qué fuerza tiene la objeción que

continuamente recibe de que el existencialismo no nos dice nada de la naturaleza, o de los valores eternos, o de Dios, y solo se dedica a hablar del hombre. Para el existencialismo, la filosofía no puede ni debe hablar de nada más. Por lo que hace a la naturaleza, las ciencias nos darán todo el control que necesitemos, y la filosofía ni puede aportarles nada ni tomar nada de ellas. No es el fundamento de la ciencia, ni los objetos de la ciencia son los suyos. Por lo que a los valores toca, los existencialistas insinúan que toda la idea de buscarles fundamentos objetivos es un timo, y que la historia de esa búsqueda no ha sido más que la de encontrar formas de defender valores que uno ya había adoptado, no de encontrar valores a los que deba plegarse. En cuanto a Dios, los existencialistas sugieren que lo único claro es que no es por investigación filosófica como se llega a él.

De manera parecida, el existencialismo no le ve la fuerza a la objeción de que el tipo de cosas que dice sobre el hombre son ambiguas, vagas, difíciles de comprender e imposibles de verificar. Para el existencialismo, no hay razón para esperar que el hombre deba ser de tal modo que de él se puedan decir cosas claras, inequívocas y fácilmente inteligibles. Es más, hay toda clase de razones para esperar lo contrario. Para el existencialismo, de hecho, lo importante del hombre, lo que hace falta decir sobre él, es que es inseguro, vago, difícil de comprender, y su camino, imposible de trazar. El problema que afrontan los filósofos existencialistas es hallar el modo de decir esto, una forma de decirlo que no haga que el público se vaya a casa diciéndose: «ah, sí, qué gran verdad. El hombre es inseguro y ambiguo. Por lo menos sabemos lo que es el hombre. Muchas gracias a los filósofos por habernos resuelto este problema».

Por tanto, y en mi opinión, lo que más distingue al existencialismo es que se las ha arreglado para desafiar una serie de supuestos que los filósofos y su público han venido aceptando inconscientemente desde Platón. Lo que lo diferencia de la filosofía tradicional no es simplemente que tome sus categorías de regiones inesperadas de la experiencia, o que se exprese de forma rara o paradójica. Todo eso ya se había hecho antes. Lo novedoso es que elige esas categorías y

se expresa de esa forma en pro de una meta nueva. Esa meta —preservar la subjetividad, negarse a buscar seguridad a base de perseguir la objetividad— es genuinamente ajena a la tradición filosófica. En lugar de intentar darle seguridad al hombre, el filosofar existencialista trata de evitar que el hombre compre una seguridad barata. En vez de explicación reductiva, el existencialista ofrece simple descripción nada más. Sus categorías —vergüenza, angustia y demás— no pretenden servir de cajones, sino de punteros. No son herramientas para lograr controlar el entorno, sino para mantenerse uno abierto al entorno. Para el existencialista, la meta de la filosofía no es hacer innecesario pensar, sino resistir todas las tentaciones de dejar de pensar, y destacadamente la tentación de la «objetividad».

Permítanme que termine esta versión enlatada del existencialismo volviendo a la cuestión que descarté al principio por descaminada: la de si es realmente una filosofía o no. Se ha aducido que, cuando uno abandona la búsqueda de objetividad, abandona la filosofía misma, y que por tanto el existencialismo no tiene derecho a reclamar ese título. Lo primero que hay que decir de este argumento es que es verbal, y, en la forma que le he dado, no merece la pena ocuparse de él. Si uno quiere limitar la palabra «filosofía» a las actividades de la gente que acepta la idea tradicional de objetividad, pues muy bien. Usar una palabra más bien que otra no cambia nada. Pero lo segundo que hay que decir es que el argumento insinúa otro problema, y ese no puede despacharse como verbal. El problema gira en torno a una idea que expresé en la primera parte de mi exposición: la de que el filósofo siempre tiene que desempeñar dos papeles, y que, en su papel de historiador de la filosofía, tiene el deber de mantener un sentido de la historia. Si algo significa ese deber, es que no debe permitirse perder la comunicación con el pasado, que ha de tener suficiente respeto por la tradición como para no dejar que haya nunca una ruptura absoluta de comunicación entre pasado y presente, que debe cultivar a propósito la distancia suficiente con sus propios compromisos para que ello le deje ver la filosofía del presente como una respuesta a las mismas preguntas que contestaba la filosofía del pasado. No se puede negar que los existencialistas cuestionan este supuesto y que, por tanto, ponen en

duda la posibilidad de que exista algo como una historia contemplativa de la filosofía. Con ello, ponen en duda la legitimidad de cualquier distinción entre el filósofo como participante y el filósofo como espectador. Ahora bien, no tengo un modo rápido de resolver esa duda. La he planteado solo para poder concluir con la siguiente pregunta, que a mi modo de ver expresa, por así decir, el atolladero existencial del existencialismo como movimiento filosófico. La pregunta es: ¿puede uno revolverse contra los ideales de seguridad y objetividad, en nombre de la inseguridad que es fundamental a la condición humana, sin perder con ello lo que sea que hace del sentido histórico algo precioso?

---

[1] Por supuesto, esta visión de la filosofía tradicional como un juego no es patrimonio exclusivo de los existencialistas. Es común a existencialistas, marxistas y filósofos analíticos. Pero de momento son los existencialistas quienes están haciendo un uso más eficaz de esa finta polémica.

[2] De hecho, a Hegel le fue tan bien en este sentido que, cuando murió, iba camino de convertirse en una poderosa influencia en asuntos políticos.

[3] Vistas las cosas desde este ángulo, la actual revuelta de los existencialistas contra la filosofía tradicional es análoga a la revuelta contra el clero durante la Reforma. En ambas, el individuo intenta borrar al intermediario y llevar en persona los necesarios tratos con las cosas elevadas.

[4] William Shakespeare, *Bien está lo que bien acaba*, acto II, escena III, 1-3. —Eds.

[5] De hecho, es una verdad trivial si con «los filósofos» se quiere decir quienes pasan su vida estudiando la historia de la disciplina, pues el filosofar nunca se ha dejado ni se dejará a tales personas.

CAPÍTULO 3
KANT COMO FILÓSOFO CRÍTICO

Esta tarde quiero presentarles con este trabajo una interpretación de Kant. Kant es la clase de filósofo al que hace falta interpretar. No basta con estudiarlo. No es ya que escribiera de forma oscura, sino que sus obras parecen apuntar a la vez en dos direcciones opuestas. En eso es como Aristóteles. La mayoría de los filósofos tienen solo un oponente; Kant y Aristóteles tenían dos. Ambos estaban reaccionando a una pareja de posiciones extremas (el dogmatismo y el escepticismo): Platón y Demócrito en un caso, y Leibniz y Hume en el otro. Dado que critican tanto al miembro dogmático de cada pareja como al escéptico, sus obras tienden a dividirse en dos conjuntos de pasajes según a quién estén criticando. La fuerza de las críticas a cada uno de sus respectivos predecesores es tal que, si prestamos atención únicamente a la parte de las obras en que están criticando a uno, es fácil pensar que son solo una versión modificada del otro. Así, si lees ciertas partes de Aristóteles nada más, parece un platónico defendiendo a Platón de las herejías de atomistas, fisicistas y sensualistas. Si solo lees otras, parece un fisicista y fenomenista democríteo desacreditando a Platón. Si lees todo Aristóteles, es dificilísimo encontrar un único cuadro coherente. Por eso existen tantos Aristóteles como especialistas aristotélicos.

Algo similar ocurre con Kant. Si solo lees la «Analítica»[1] o la primera parte de los *Prolegómenos*, saltándote un poco aquí y allá, es fácil pensar que sale en auxilio de los racionalistas contra Hume. Si lees la «Dialéctica trascendental» nada más, o solo la última parte de los *Prolegómenos*, resulta fácil verlo como un positivista lógico nacido antes de hora, ofreciendo una teoría verificacionista del significado con cuya ayuda poder desacreditar a los metafísicos. Según qué parte del libro tomes como núcleo de la *Crítica de la razón pura*, puedes hacer de Kant un racionalista particularmente eficaz o un empirista particularmente eficaz. Pero cuando intentas

formular «doctrinas kantianas» diferenciables y positivas (doctrinas que sean su propia contribución positiva), lo pasas francamente mal. Por eso existen tantos Kant como kantianos, mientras que no hay tantos Hume como humeanos ni tantos Leibniz como leibnizianos.

Por expresar el problema de otra manera, cuando un filósofo es tan profundo y tan original como Aristóteles o Kant, parece que cuesta mucho concebirlo como alguien que hace solo trabajo negativo, que solo desnuda los errores de sus predecesores extremistas. Con todo, es sumamente difícil hallar en la *Metafísica* de Aristóteles o en la primera *Crítica* de Kant otra cosa que argumentos negativos. Cuando uno encuentra argumentos positivos, parecen la parte más floja del libro. Por ejemplo, la única sección claramente positiva y original de la *Metafísica* es la prueba de la existencia del Motor Inmóvil, y ese es con seguridad el argumento más dudoso del libro. La sección de la *Crítica* que parece más positiva y original es la teoría de las sucesivas síntesis de lo múltiple que encontramos en la «Deducción de las Categorías» y sus alrededores, pero también ese parece el tramo más dudoso de la obra. Además, estas secciones positivas son justo la parte de ambas obras que más sujeta está a interpretaciones divergentes. Nadie sabe muy bien qué entender en la noción de Motor Inmóvil de Aristóteles, pero todo el mundo entiende algo. Nadie tiene demasiado claro cómo describir lo que se propone Kant con la «Deducción Trascendental», pero todo el mundo tiene la impresión de que, tal como él la da, su propia descripción de lo que está haciendo no vale.

Resumiendo este primer punto, es mucho más fácil saber en contra de qué estaban Kant y Aristóteles, que saber de qué estaban a favor; es más, incluso las críticas a cada uno de sus respectivos oponentes son muy difíciles de formular. Cuando uno lo intenta, se ve diciendo cosas que no pueden conciliarse con las críticas al *otro* oponente respectivo.

Creo que la explicación de esto es que Kant y Aristóteles ocupan una posición casi única en la historia de la filosofía. Son filósofos *críticos*. Su logro está en sus negaciones, no en cualquiera de las tesis positivas que propusieron. Su grandeza no radica en la importancia de lo que descubrieron, sino en la profundidad de sus

críticas a otros filósofos. Son los grandes entrometidos de la filosofía. Mientras que Platón, Descartes, Spinoza, Hegel y Whitehead son grandes sintetizadores dogmáticos, y Demócrito, Hobbes, Hume y Ayer grandes analistas e impugnadores escépticos, Aristóteles y Kant se nos presentan con filosofías críticas. En eso se parecen a un tercer gran filósofo, Wittgenstein. Como Aristóteles y Kant, él no demuestra ninguna tesis ni descubre nada, todo lo que hace es mostrar a otros filósofos dónde se equivocaron. Para Wittgenstein, el objetivo de filosofar no es hacer descubrimientos o proporcionar fundamentos a lo que ya sabemos, sino sencillamente *entrometerse*. Aquí eso significa entrometerse en lo que otros filósofos han hecho enseñándoles dónde se armaron un lío, cómo decir lo que quieren decir sin despeñarse en extremos, y que los problemas que se preocupan de resolver no tienen en realidad mucho de problemático. Mostrarles, como dijo Berkeley, que han estado «levantando polvo y quejándose después de que no ven nada».

Quiero esbozar qué aspecto tiene Kant si interpretamos que hacía filosofía como la hacía Wittgenstein, como un filósofo crítico *meramente*. No pretendo que el Kant que voy a presentarles sea el Kant-en-sí: Kant tal como le parecía a Kant. Al contrario, diré continua y descaradamente que Kant interpretó mal su propia obra, que juzgó mal su propio logro y que describió mal su propio método. A cambio, el Kant que voy a presentarles es, por lo menos, un filósofo interesante. (Al menos, yo creo que lo es.)

Entendida como yo quiero entenderla, filosofía «crítica» significa el tipo de filosofía que acepta ciertos datos como sencillamente dados. Ni siembra dudas sobre ellos, ni intenta justificarlos. Para Kant, esos datos eran la ciencia y la matemática de su tiempo y la conciencia moral común de su tiempo. Una filosofía crítica no pregunta si sus datos son reales; presupone que lo son. En eso difiere de una filosofía escéptica. Pero tampoco intenta proporcionarles un *fundamento*. En eso difiere de una filosofía dogmática. Kant, desde mi punto de vista, simplemente no tiene nada que decir en respuesta a quien no cree que las verdades sintéticas a priori que enumera en la *Crítica* (los «Principios del Entendimiento Puro») son ciertas, aunque tiene mucho que decir

contra los argumentos escépticos según los cuales no podemos saber que lo son. Análogamente, no tiene argumentos que ofrecer contra los filósofos que sencillamente no pueden encontrar en sí mismos ese sentido del deber que, dice Kant, es el contenido de la conciencia moral común. Por lo tanto, no tiene argumentos que dar a Trasímaco, o a Nietzsche, o a la ética del caballero de Aristóteles. Si no te parece verdadero el principio de causalidad, o el imperativo categórico, Kant no te los va a demostrar. Tal como yo lo veo, él piensa en la verdad de los Principios del Entendimiento Puro (por ejemplo, el principio de causalidad) y en la del imperativo categórico como datos que tocan fondo absolutamente. No cabe meterles un fundamento por debajo ni demostrar que son falsos. Por consiguiente, para Kant los problemas filosóficos no son de la forma «¿por qué son verdaderos esos principios?» Más bien, su forma es «¿por qué ha pensado nadie alguna vez (en particular, los filósofos escépticos) que eran falsos?», o «¿por qué ha pensado nadie que necesitaban fundamento (como han hecho los filósofos dogmáticos)?»

Por tanto, es un error quejarse demasiado de que Kant nunca pruebe realmente que sus doce categorías son una lista exhaustiva de los distintos modos de sintetizar la multiplicidad de la intuición. También es un error protestar por que sus formulaciones del imperativo categórico no sean de ninguna ayuda al intentar resolver dilemas morales reales. Es totalmente cierto que no demuestra que esa lista sea la lista de categorías «correcta», ni consigue darte un método a toda prueba para decidir casos de conflicto entre deberes. Pero eso no es muy importante, incluso si Kant hubiera pensado que lo era, que seguramente sí. Lo importante de Kant es la forma en que muestra cómo es posible que haya verdades sintéticas a priori, no qué verdades concretas pensaba él que eran sintéticas a priori, y que es posible que haya un imperativo categórico, no qué enunciados formulan con propiedad ese imperativo. El interés y el valor de Kant como filósofo seguirían siendo los mismos incluso si dejáramos de creer en cualquiera de los doce Principios del Entendimiento Puro e incluso si rechazáramos del todo su descripción de la conciencia moral común. Seguirían siendo los

mismos porque su obra entera puede tomarse como una proposición hipotética: dando por sentado que estos datos están dados, ¿cómo son posibles? Dicho de otro modo, dando por sentado que estos son los datos que hallamos, ¿por qué los filósofos no pueden dejarlos estar? ¿Por qué, en otras palabras, presentan continuamente cargos escépticos contra estos datos e intentan socavarlos (como Hume), o bien tratan de impedir que sean socavados y, por tanto, intentan meterles debajo fundamentos (como Leibniz)? Como filósofo *crítico*, esa es la pregunta que Kant está respondiendo. Su problema no es qué creer o qué hacer, sino meramente por qué los filósofos han dicho lo que han dicho. Mientras que los filósofos escépticos niegan que los datos sean como son, y los dogmáticos piensan que necesitamos descubrir nuevos datos para asegurarnos de los originales, Kant simplemente da por hechos los datos y pregunta por qué esos dos tipos de filósofo dicen las tonterías que dicen.

Tengo que detenerme un momento para aclarar una confusión. Podría parecer natural sugerir que la pregunta «¿cómo es posible tal y tal dato?» (por ejemplo, «¿cómo son posibles los juicios sintéticos a priori?» o «¿cómo es posible un imperativo categórico?») es la petición de un fundamento, de una explicación causal, como han insinuado los filósofos dogmáticos. Por ejemplo, puede parecer que la pregunta «¿cómo es posible el conocimiento matemático?» exige una respuesta causal: o bien una que venga dada por causas directamente físicas, en términos de fisiología cerebral, o por causas metafísicas, como en la doctrina de Platón de las Ideas y/o de la Reminiscencia. Puede parecer que la pregunta «¿cómo es posible la moralidad?» tiene que ser la petición de explicar causalmente de dónde viene la ley moral, de modo que podamos recurrir a explicaciones psicológicas y antropológicas, por un lado, o a explicaciones metafísicas en términos de una reminiscencia de la Forma del Bien o de la implantación de un «sentido moral» por parte de un ser divino. Pero, en el sentido en que Kant plantea la pregunta «¿cómo es posible el dato?», no se trata de pedir una explicación causal. De hecho, la contribución singular más grande que hizo Kant a la filosofía fue ayudarnos a ver que el tipo de explicaciones que esta ofrece, en oposición a la ciencia, no son causales. Para un

filósofo crítico, preguntar «¿cómo es posible el dato?» es simplemente un medio para suscitar otra pregunta, «¿qué te hace dudar de que sea posible?» Lo que quiere hacer es desnudar las razones que los filósofos han dado para dudar de la posibilidad del dato. Normalmente, esas razones las han formulado primero filósofos escépticos. Pero luego las han adoptado los filósofos dogmáticos, que le tienen tanto miedo al escéptico que creen que tienen que proporcionarle al dato un fundamento. El cometido principal de la filosofía crítica consiste en despejar las razones del escéptico para dudar o las razones del dogmático para alcanzar un «fundamento».

En Kant, concretamente, el cometido principal de la *Crítica de la razón pura* es desnudar el mito escéptico humeano de que podemos analizar el conocimiento en términos de intuiciones (lo que Hume llamaba «impresiones») sin conceptos, y el mito racionalista leibniziano de que podemos analizar el conocimiento en términos de conceptos puros sin intuiciones sensoriales. Una vez que quedan en evidencia esos mitos, se nos quita el ansia de dar explicaciones causales de nuestras bases; somos capaces de aceptarlas como lo que son, datos que tocan fondo. Este dejar al desnudo es a lo que antes me refería con el *entrometerse* en que consiste el tipo especial de filosofía practicada por Aristóteles, Kant y Wittgenstein. Es también lo que tenía en mente cuando dije que lo que resultaba interesante e importante en sus obras no eran sus tesis positivas, sino las críticas negativas a sus predecesores. La clase de respuestas que da Kant a la pregunta «¿cómo es posible el dato?» son de la forma «solo parece imposible o dudoso si lo miras con las gafas equivocadas». Quítate las gafas y dejará de parecerte imposible o dudoso. Más exactamente, solo parece imposible o dudoso si hablas de él en el lenguaje equivocado. Usa un lenguaje diferente y podrás tomarlo por lo que dice. En suma, para un filósofo crítico los problemas filosóficos no se resuelven; más bien (por utilizar la expresión ahora de moda), se «disuelven». No contestamos a la pregunta «¿cómo es posible el dato?» a la manera tradicional y directa de proveer una explicación causal de él, sino en la forma oblicua de explicar por qué lo preguntan los que hacen la pregunta,

y eliminando a continuación su razón para hacerla. Explicar cómo es posible el conocimiento sintético a priori no es como explicar cómo es posible el arco iris (donde damos una explicación causal). Se parece más a explicarle a un niño cómo es posible que la palabra «río» sea a la vez el nombre de una corriente de agua y un verbo que significa «lanzo una risa».[2] Al tratar con el niño no damos una explicación etimológica del origen de las dos palabras, y si lo hiciéramos no le dejaría satisfecho. En vez de eso, intentamos eliminar su falsa suposición sobre la naturaleza del lenguaje, a saber, la de que un mismo sonido solo puede estar por una única palabra.

La idea sobre los métodos y el propósito crítico de la filosofía que estoy recomendando como ayuda para interpretar a Kant se puede resumir en varias citas de Wittgenstein. Estos pasajes se reproducen bastante a menudo, pero creo que lo que dice en ellos es importante y no viene mal repetirlo:

(1) Toda *explicación* tiene que desaparecer y solo la descripción ha de ocupar su lugar. Y esta descripción recibe su luz, esto es, su finalidad, de los problemas filosóficos. Estos no son ciertamente empíricos, sino que se resuelven mediante una cala en el funcionamiento de nuestro lenguaje, y justamente de manera que este se reconozca: *a pesar de* una inclinación a malentenderlo.[3]

(2) La filosofía expone meramente todo y no explica ni deduce nada. —Puesto que todo yace abiertamente, no hay nada que explicar. […] El trabajo del filósofo es compilar recuerdos para una finalidad determinada.[4]

(3) Mis proposiciones esclarecen así: quien me entiende las reconoce al final como absurdas, cuando a través de ellas —sobre ellas— ha salido fuera de ellas. (Tiene, por así decirlo, que arrojar la escalera después de haber subido por ella.)[5]

En lo que sigue, quiero aplicar a Kant esta idea de lo que la filosofía puede y debería ser. Creo poder mostrar, por lo menos, que si se usa para interpretar qué pretende Kant, mucho de lo que hay

en él de valioso se puede preservar y muchas dificultades que surgen al interpretarle pueden resolverse. Es posible que, si uno piensa en el logro filosófico de Kant en estos términos, pierda de vista elementos importantes de su pensamiento. En realidad, en mi opinión no es así, pero esa es una de las cosas que podemos discutir después.

LA TEORÍA DEL CONOCIMIENTO DE KANT. LA REVOLUCIÓN COPERNICANA

Me dejo ahora de generalidades sobre qué tipo de filósofo era Kant y paso a intentar decir algo concreto sobre su tesis central en teoría del conocimiento: la de que, como suponer que todo nuestro conocimiento debe conformarse a los objetos hace que no podamos responder a la pregunta «¿cómo es posible el conocimiento sintético a priori?», tenemos que adoptar el supuesto alternativo de que «los objetos deben conformarse a nuestro conocimiento». Hay varias maneras de interpretar esa última frase, «los objetos deben conformarse a nuestro conocimiento». En primer lugar, podemos tomarla al pie de la letra, en cuyo caso es falsa obviamente y sin remedio. Porque, si algo hay claro en el conocimiento humano, es que no cambia las cosas que conocemos. Si uno no se mueve de aquí, entonces no hay más pregunta que la de cómo pudo pensar Kant jamás que hacía falta decir algo tan raro. La respuesta parece bastante fácil. El motivo por el que pensó que tenía que decirlo era que había aceptado la teoría cartesiana de las ideas como copias y había descubierto que solo conducía al escepticismo o al dogmatismo. No veía ningún otro modo de escapar a ambos que obligarse a abandonar la teoría de las ideas como copias, o sea, la suposición de que una idea es una suerte de réplica inmaterial de un objeto material o de una parte de él. Renunció a pensar que nuestras ideas copiaban lo real y adoptó la única alternativa que se le ocurrió: a saber, que nuestras ideas *hacían* el mundo. Pero, por desgracia, esa alternativa resultaba absurda.

Lo que acabo de esbozar es una interpretación de Kant bastante extendida. Permítanme que la describa con un poco más de detalle. La teoría de las ideas como copias era común a todos los filósofos desde Descartes hasta Hume. De acuerdo con ella, conocemos la verdad solo cuando nuestras ideas copian exactamente las cosas. Pero, tal como Descartes y Berkeley dejaron claro, nunca estamos en situación de comparar nuestras ideas con las cosas. No puedes saltar fuera de tu mente y ver si lo que hay en ella encaja con lo que hay afuera. Por consiguiente, parece que tenemos que elegir entre el escepticismo de Hume y una garantía metafísica de que algunas de nuestras ideas realmente sí copian exactamente las cosas. Es esa clase de garantía metafísica la que proporcionaban el Dios de Descartes, la teoría de Spinoza de los dos atributos paralelos y la armonía preestablecida de Leibniz. Tales garantías metafísicas son lo que Kant considera productos típicos de la metafísica dogmática. Si tu estómago no tolera el escepticismo de Hume ni la metafísica dogmática, como le pasaba a Kant, entonces tienes que abandonar la teoría de las ideas como copias. Si abandonas esa teoría, necesitas otra cosa que ocupe su lugar. Como Kant no pudo pensar en nada salvo una simple inversión —si las ideas no pueden conformarse a las cosas, las cosas deben conformarse a las ideas—, nos salió con la palmaria ridiculez de que nuestras ideas tienen el poder de afectar, o hasta producir, cosas físicas reales.

Si uno interpreta la Revolución Copernicana como acabo de proponer, entonces deberá concluir que, aunque Hume, Leibniz y más o menos todos los demás estaban equivocados sin remedio sobre la naturaleza de nuestras ideas, y que esa Revolución hacía muchísima falta, la de Kant no vale para nada. Si se sigue hasta el final esta línea de pensamiento, se llega a las mismas conclusiones que aristotélicos, tomistas y heideggerianos: a saber, que hay que abandonar tanto la teoría de las ideas como copias cuanto la absurda alternativa de Kant a ella. Por tanto, se tiene que adoptar una teoría nueva sobre la naturaleza de las ideas. Esa nueva teoría acaba siendo normalmente una vieja, atribuida a Aristóteles. Es la de que las ideas son entidades que poseen un tipo único de existencia llamada «existencia intencional», que tiene la peculiar propiedad de

que las ideas pueden ser al mismo tiempo idénticas a las cosas y diferentes de ellas. La propiedad me parece verdaderamente peculiarísima, y, a la postre, demasiado peculiar como para servir de fundamento a una teoría epistemológica que funcione. De todos modos, no quiero discutirla aquí. Solo quiero observar que la línea argumental que he venido señalando es la que lleva a la gente a decir que el saldo neto de la obra de Kant es dejar clara la bancarrota del cartesianismo al poner al descubierto el absurdo de la teoría de la copia, y que, por tanto, su moraleja es *¡vuelta a los griegos!* (Y, en concreto, *¡vuelta a Aristóteles!*).

Por mi parte, estaría completamente de acuerdo en que Kant demostró la bancarrota del cartesianismo, pero sostendría que la moraleja que hay que extraer no es *¡vuelta a los griegos!*, sino, dicho rápidamente, *¡adelante hasta Wittgenstein y el análisis lingüístico!* Déjenme sugerir ahora una segunda forma en que se puede interpretar la expresión «los objetos deben conformarse a nuestro conocimiento». En vez de pensar que Kant dijo (de manera absurda) que las cosas cambian por nuestro conocimiento de ellas, interpretemos que dice que el conocimiento es el producto de dos procesos causales: uno que opera desde el lado del yo y otro desde el lado del objeto. En otras palabras, interpretémosle como si dijera que, aunque las cosas-en-sí por supuesto que no cambian por nuestro conocimiento de ellas, el «objetos» en la expresión «los objetos deben conformarse a nuestro conocimiento» no significa «cosas-en-sí». Más bien, ahí significa los productos de la transacción causal entre yoes-en-sí y cosas-en-sí. Esos productos son los objetos espaciotemporales corrientes —las mesas y las sillas— de la vida cotidiana. Desde este punto de vista, podemos interpretar que Kant dice que las mesas y sillas, y los objetos físicos en general, se hacen tomando alguna materia de las cosas-en-sí (lo «dado puro») y algunas formas del yo-en-sí (el espacio, el tiempo, las categorías y algunos conceptos empíricos) y juntándolas de algún modo complicado. Si seguimos esta interpretación, entonces tiene perfecto sentido decir que «los objetos deben conformarse a nuestro conocimiento», pues en este caso todo lo que estamos diciendo es que solo podemos ser conscientes de la cosa-en-sí en la medida en

que nuestras ideas ya le hayan dado forma. Con esta interpretación, podemos ver ahora que la respuesta a la pregunta «¿cómo es posible el conocimiento sintético a priori?» es simplemente que podemos saber por adelantado qué aspecto van a tener las cosas porque, si no tuvieran ese aspecto, nunca seríamos capaces de hacernos conscientes de ellas.

Para empezar, pienso que esta interpretación de la Revolución Copernicana se aproxima bastante a lo que Kant tenía en mente. Además, no creo que la Revolución Copernicana tenga nada de malo así interpretada, siempre que uno se abstenga de preguntar por la naturaleza de la cosa-en-sí y del yo-en-sí. Dicho de otro modo, la interpretación funciona mientras no hagas preguntas engorrosas sobre cómo es que Kant sabe tanto del proceso de constituir objetos espaciotemporales a partir de componentes no-espaciotemporales. En cuanto empiezas a hacer esas preguntas, esta interpretación de lo que está haciendo Kant se vuelve incoherente. Seguramente las razones les son conocidas a ustedes, pero probaré a resumirlas.

La tesis principal de la «Analítica Trascendental» es que todos nuestros conceptos son herramientas para captar intuiciones presentadas espaciotemporalmente. Más en concreto, lo que resulta de la «Deducción Trascendental de las Categorías» no es solo que «los pensamientos sin intuiciones son vacíos [y] las intuiciones sin conceptos son ciegas».[6] Lo que aflora es la tesis más fuerte de que es imposible ser consciente de un concepto sin usarlo para sintetizar algunas intuiciones; y no es posible ser consciente de una intuición sin haberla ya sintetizado con otras con la ayuda de conceptos. Esto significa que cualquier conocimiento que tengamos debe ser un caso de síntesis bajo conceptos de una multiplicidad de intuiciones. Puesto que todas las intuiciones deben ser espaciotemporales, eso quiere decir que cualquier conocimiento que tengamos *debe ser conocimiento de objetos espaciotemporales*. Si lo que dice Kant en la «Analítica» es correcto, es absolutamente imposible tener conocimiento de ninguna otra cosa. Si Kant tiene razón, hay solo tres clases de proposiciones que puedan llamarse verdaderas: las proposiciones analíticas, las proposiciones necesarias sobre objetos

espaciotemporales (las verdades sintéticas a priori) y las proposiciones contingentes sobre objetos espaciotemporales (las verdades sintéticas a posteriori).

Pero, si es así, ¿cuál es el estatus de las proposiciones que Kant asevera en la *Crítica* (como, por ejemplo, que la cosa-en-sí es la fuente de las intuiciones, o que en la actividad del yo-en-sí de unificar la multiplicidad de intuiciones hay involucradas tres síntesis)? ¿Se trata de actividades realizadas por la sustancia espaciotemporal y que pueden disponerse en secuencias causales espaciotemporales? Es obvio que no. El propio Kant diferencia firmemente entre el yo empírico y el cuasi-yo, al que llama «la unidad trascendental de apercepción» o el «ego trascendental». El yo empírico es temporal, e incluso espacial en cierto sentido, y sus actividades se pueden observar y describir más o menos del mismo modo que observamos y describimos el comportamiento de mesas, sillas y conejos. Pero la unidad trascendental de apercepción no es, y es imposible que sea, una sustancia espacial o un agente causal. Más obvio todavía, cualquier cosa que se diga sobre la cosa-en-sí es automáticamente una proposición sobre lo no-espaciotemporal. Es, por tanto, una proposición de la que es imposible que sepamos que es verdadera, sobre la base del propio Kant.

Por lo tanto, parece razonable concluir que, si interpretamos la expresión «los objetos deben conformarse a nuestro conocimiento» en términos de una serie de enunciados sobre una transacción causal entre el yo-en-sí y la cosa-en-sí, entonces Kant se está contradiciendo de modo flagrante. Prácticamente nada de lo que dice en la «Analítica» sobre la forma en que nuestros conceptos constituyen los objetos que conocemos *puede ser dicho* si lo que dice sobre las limitaciones de nuestro conocimiento es verdad.

Alcanzada esta conclusión, uno puede seguir varias vías. En primer lugar, puede decir que es posible reinterpretar lo que está diciendo Kant en la «Analítica trascendental» de modo que todo se convierta en una serie de proposiciones psicológicas sobre las actividades del yo empírico. En otras palabras, intentar interpretar la discusión en torno al papel de las intuiciones y los conceptos en nuestro pensamiento como una serie de descubrimientos empíricos

sobre secuencias temporales de sucesos mentales que pueden observarse por introspección. En mi opinión, esta alternativa es totalmente inviable y no seguiré con ella. Se reduce al intento de transformar a Kant en un simple empirista humeano más. Aunque yo mismo intentaré transformarle en una especie de analista lingüístico wittgensteiniano, creo que la transformación humeana se pasa de la raya.

Una segunda vía sería decir que la «Analítica» no nos está hablando de entidades espaciotemporales observables empíricamente, sino de un tipo especial de entidad espaciotemporal. Visto así, Kant es un tipo peculiar de metafísico. Si uno elige esta línea, dirá que las proposiciones que describen la actividad del yo para dar forma a la materia suministrada por la cosa-en-sí son de una clase muy especial, que exhiben una clase muy especial de conocimiento que se llama conocimiento «trascendental». Ese conocimiento «trascendental» será algo a medio camino entre nuestro conocimiento científico corriente de entidades espaciotemporales y el conocimiento metafísico «trascendente» del que Kant se burla (en la «Dialéctica Trascendental») cuando se lo atribuyen los filósofos dogmáticos. En otras palabras, uno tendrá que decir que, en la «Analítica Trascendental», Kant nos está dando un tipo de conocimiento que escapa de alguna forma a las restricciones que está proponiendo sobre el alcance del conocimiento humano, pero que resulta necesario para establecer esas restricciones. Una forma alternativa en que puede describirse esta postura es pensar que lo que Kant nos ofrece es una «metafísica del ego trascendental». También esta opción me parece completamente estéril. Se reduce a intentar sacar a Kant del agujero a base de postular un novísimo tema de indagación metafísica, un novísimo tipo de metafísica, y un novísimo tipo de conocimiento metafísico llamado conocimiento «trascendental». Esta táctica no soluciona ninguno de los problemas. Simplemente los bautiza.

Yo pienso que, para sacar a Kant del agujero, la opción que uno intente tiene que ser otra y completamente diferente. Hay que retroceder y preguntar una vez más por el sentido de la expresión «los objetos deben conformarse a nuestro conocimiento».

Necesitamos ver si se puede rescatar de ella algún significado que no haga aterrizar a Kant en el absurdo o en la autocontradicción. Consideremos por un momento la expresión aislada. Kant está hablando aquí como si coincidiera con la tradición cartesiana en que hay dos clases diferenciables de entidades: por un lado, los actos de conocer humanos, y, por otro, las cosas que resultan conocidas por los hombres. Dicho de otra manera, está hablando como si hubiera dos tipos de cosas descriptibles en los mismos términos y el único problema fuera cuál debe conformarse a cuál. Ahora preguntemos: ¿en qué términos podemos describir ambos conjuntos de entidades? ¿Cómo hallaremos un modo neutral de describir tanto el conocimiento humano como lo que el conocimiento humano conoce? Quiero proponer que la moraleja de la «Analítica Trascendental» es que no tenemos modo de hacerlo. O sea, quiero proponer que lo que Kant nos muestra en la «Analítica» es que la entera enunciación del problema —¿las ideas se conforman a los objetos, o los objetos a las ideas?— es totalmente confundente. Las preguntas sobre qué se conforma a qué tienen sentido cuando estamos hablando de objetos espaciotemporales. En esos casos, puede ser sensato preguntar si se parecen uno a otro y cabe preguntar cuál influyó de qué maneras sobre cuál. No tienen ningún sentido si lo que se nos pide es que comparemos una entidad espaciotemporal y otra no-espaciotemporal y digamos cuál influye en cuál, y si coinciden o hasta qué punto lo hacen. La «Analítica» muestra que todos nuestros conceptos, y por tanto, también todas nuestras maneras de hablar sobre objetos de investigación, están diseñados para hablar sobre objetos espaciotemporales, sobre sustancias que entran en interacción causal. Pero si, como muestra la «Analítica», los conceptos y las intuiciones no son en absoluto sustancias espaciotemporales, entonces necesitamos un nuevo conjunto de términos. Por eso creo que, si intentamos darle sentido a la enunciación original del problema por parte de Kant, expresada como está en términos tradicionales, inevitablemente fracasaremos; y que, por consiguiente, debemos buscar una forma radicalmente nueva de enunciar el problema y de enunciar la así llamada Revolución Copernicana. Kant enunció el problema de la epistemología «¿cómo

son posibles los juicios sintéticos a priori?» en unos términos que insinúan que lo que se requería era una explicación causal. Su respuesta, «son posibles porque los objetos deben conformarse a nuestras ideas», parece suministrarla. La analogía con Copérnico confirma esta insinuación de que era una explicación causal lo que se pedía y lo que se dio. El problema de Copérnico era «¿por qué los cuerpos celestes se mueven como se mueven?», y su respuesta fue proponer una nueva pauta de interacción causal entre los órganos sensoriales de los hombres en la Tierra y esos cuerpos. La enunciación que da Kant a su problema y su solución a él hacen pensar que también él está proponiendo una pauta nueva de interacción causal entre dos entidades del tipo de las que pueden establecer entre sí interacciones causales, a saber, del tipo de las sustancias espaciotemporales.

Pero, si lo que Kant dice en la «Analítica» está bien, ya no podemos seguir pidiendo respuestas a los problemas epistemológicos como si las estuviéramos pidiendo para problemas científicos. Y, por consiguiente, no podemos seguir dando explicaciones epistemológicas como si fueran explicaciones causales. Así que tenemos que concluir que, al escribir el prefacio a la segunda edición (donde figura el pasaje sobre la Revolución Copernicana), Kant no vio el impacto que tenía su propio análisis sobre la descripción que él mismo da del problema que ese análisis está destinado a resolver. Esto significa que debemos encontrar un nuevo modo de describir y enunciar los problemas epistemológicos. Supongamos que probamos con lo que sugerí antes en esta charla e interpretamos la pregunta «¿cómo son posibles los juicios sintéticos a priori?» como «¿por qué los filósofos han dudado de su posibilidad?» Pongamos que lo hacemos. ¿Cómo podemos interpretar, entonces, la respuesta de Kant? Propongo que lo hagamos así: los filósofos han dudado de que los juicios sintéticos a priori sean posibles porque han descrito las entidades mentales de un modo que da a entender que cualquier intuición dada puede existir independientemente de cualquier concepto dado, y viceversa. Es decir, han concebido los conceptos y las intuiciones de tal manera que cada concepto podría ser un contenido de conciencia sin que

necesitemos ser conscientes de ninguna intuición. Y, a la inversa, que se puede tener cualquier intuición en la conciencia sin ser necesariamente consciente de ningún concepto. El error fundamental encerrado aquí es que los filósofos han descrito los conceptos y las intuiciones como describimos las mesas y las sillas, como diferentes cuasi-sustancias que existen codo con codo en un cuasi-espacio mental e interactúan unas con otras de forma cuasi-causal. Por tanto, han pensado que una intuición es tan buen candidato a ser un «contenido de conciencia» como una mesa o una silla. Dada esa imagen de la mente, no sorprende que aparezcan dudas sobre las proposiciones sintéticas a priori. Porque son proposiciones que vienen formuladas exclusivamente en términos de conceptos y, aun así, te dicen algo sobre las intuiciones. Como, en la visión tradicional, está claro que las intuiciones son un tipo de entidad completamente distinta de los conceptos, aunque ambas sean mentales, es difícil ver cómo *podrían* los conceptos constreñir necesaria y universalmente las intuiciones. De hecho, es imposible entender que puedan; tan imposible, y por las mismas razones, como ver posible que un cierto tipo de objeto físico deba mantener necesaria y universalmente determinadas relaciones con otro tipo de objeto físico. Una vez que se piensa en la mente como un espacio interno lleno de entidades mentales autosuficientes que interactúan, tan difícil es ver cómo podría haber conexiones *necesarias* entre esas entidades mentales, como ver cómo podría haber conexiones *necesarias* entre sustancias espaciotemporales que interactúan.

Así que la respuesta a «¿cómo podemos hacer que la gente deje de poner en duda que hay principios sintéticos a priori?» es la siguiente: lo podemos lograr si conseguimos que dejen de hablar de las entidades mentales como hablan de las espaciotemporales. Si dejan de hacerlo —concretamente, si dejan de hablar de los conceptos como si fueran un tipo de objeto de la atención introspectiva y de las intuiciones como si fueran otro—, a lo mejor ven entonces cómo son posibles las verdades sintéticas a priori, o, dicho de otro modo, a lo mejor dejan de hacer pseudo-preguntas sobre ellas. Más concretamente todavía, si ven que un concepto no es un tipo de entidad mental que pueda existir en la conciencia al

margen de su uso como sintetizador de intuiciones, y que una intuición no es un tipo de entidad mental que pueda existir en la conciencia al margen de haber sido sintetizada por conceptos, entonces verán que es perfectamente plausible que algunos conceptos se usen en la síntesis de todas las intuiciones. (Por ejemplo, el concepto de Sustancia está en ese caso.) Ya no serán incapaces de entender, por consiguiente, la posibilidad de principios sintéticos a priori.

Ahora bien, ¿dónde encaja en todo esto la Revolución Copernicana? Probablemente mi respuesta les chocará por parecerles un poco descabellada y estar muy lejos de cualquier cosa que pudiera pensar Kant. Pero la daré de todos modos, basándome en mi afirmación previa de que cualquier interpretación más convencional nos lleva a la conclusión de que Kant, o estaba diciendo algo absurdo, o se contradecía. Creo que la verdadera revolución en epistemología que trajo Kant fue hacernos ver simplemente que el lenguaje en que describimos la consciencia y el conocimiento no necesita ser el mismo con el que describimos las entidades físicas que conocemos y de las que somos conscientes. Dicho de otra manera, nos hizo ver que nuestros problemas epistemológicos no son problemas reales a los que haya que responder haciendo descubrimientos científicos o metafísicos, sino problemas verbales. Son problemas que vienen de intentar hablar de la consciencia y del conocimiento del mundo físico del mismo modo en que lo hacemos del propio mundo físico, y que, por tanto, desaparecerán si resistimos la tentación de hablar así. Cuando Kant nos dice que un concepto no es una entidad por derecho propio, sino meramente un modo particular de unificar intuiciones, y que tampoco una intuición lo es, sino que es siempre y únicamente un mero componente analítico de una multiplicidad sintetizada, está dando el primer paso en la dirección correcta. El segundo, que él no dio, pero que sí ha dado la filosofía lingüística moderna, es darse cuenta de que la propia tesis —la de que conceptos e intuiciones *no son* la misma clase de cosa que las entidades físicas y *no cabe* pensar que establecen relaciones causales entre ellos— no expresa un descubrimiento sobre la naturaleza de las entidades mentales,

sino que es una puntualización meramente sobre nuestro lenguaje y sobre las imágenes alojadas en él.

Voy a expresar de otra manera ese segundo paso haciendo y contestando la siguiente pregunta: cuando reinterpretamos del modo que propone Kant en la «Analítica Trascendental» la naturaleza de conceptos e intuiciones, ¿qué tipo de investigación estamos llevando a cabo? Sin duda, no una investigación psicológica. Sin duda, no una investigación metafísica. Sin duda, no una mera investigación lógica, en el sentido en que Russell y Whitehead la practicaban. Decir que nos ocupamos de una investigación epistemológica o de una investigación trascendental no ayuda en nada. Esos nombres no son más que etiquetas de problemas. Pero, si no es psicológica, ni empírica, ni lógica, ¿qué clase de investigación es? ¿Qué son esas cosas llamadas conceptos, intuiciones, síntesis, juicios y demás, de las que Kant no para de hablar? ¿A qué género pertenecen? ¿Qué son las facultades de Sensibilidad, Entendimiento y Razón que él distingue? Con seguridad, no son porciones diferentes del yo empírico. Así que, ¿de qué estamos hablando? Sugiero que eso de lo que estamos hablando son *palabras*. Sugiero, además, que lo que Kant está haciendo en la *Crítica* es justo lo que los filósofos lingüísticos sostienen que debe hacer la filosofía: a saber, (en primer lugar) explicar cómo usamos ciertas palabras y (en segundo lugar) proponer formas nuevas de usarlas que eviten los problemas verbales que crean las formas viejas. No sugiero ni por un momento que eso sea lo que Kant pensaba que estaba haciendo, pero sí que es la única forma de describir lo que hacía sin quedar enredado en paradojas.

Ahora tomaré otra vez toda la cuestión de la Revolución Copernicana desde el principio e intentaré hacer más plausible esta sugerencia.

LAS TESIS DE KANT COMO PROPOSICIONES ANALÍTICAS

Antes dije que, en la «Analítica», Kant había mostrado que hay solo tres clases de proposiciones que podríamos saber que son verdaderas: las analíticas, las sintéticas a priori sobre objetos espaciotemporales y las sintéticas a posteriori sobre objetos espaciotemporales. Si ahora puedo dar por sentado que, sean lo que sean los conceptos, intuiciones, síntesis, juicios y demás, no se trata de objetos espaciotemporales, resulta claro que las proposiciones que conforman las observaciones de Kant acerca del conocimiento humano no son sintéticas. Las proposiciones sintéticas son proposiciones sobre la experiencia; la experiencia lo es solo de objetos espaciotemporales; por consiguiente, no hay proposiciones sintéticas sobre intuiciones, juicios, síntesis, conceptos y demás. La única conclusión posible es que todas las observaciones de Kant sobre el conocimiento humano deben interpretarse, sobre la base del propio Kant, como proposiciones analíticas. (Esta conclusión, dicho sea de paso, parece ser aceptada por el propio Kant en algunos pasajes: por ejemplo, dice explícitamente que «el principio de la necesaria unidad de apercepción» es una proposición analítica.[7])

Si interpretamos la *Crítica de la razón pura* como un conjunto de proposiciones analíticas, ahora tenemos que indagar qué es una proposición analítica y cómo las descubrimos. La descripción del propio Kant del proceso de descubrirlas es metafórica, confundente y de escasa ayuda en general: dice que lo hacemos «descomponiendo [el concepto del sujeto] en sus conceptos parciales, los cuales eran ya pensados en dicho concepto del sujeto (aunque de forma confusa)».[8] Esto es insinuar que existe un proceso mental de descomponer conceptos análogo al proceso físico de partir una piedra o desmontar una máquina. Peor aún, que los conceptos son algo que se puede captar por introspección de forma aislada. O sea, que uno puede dirigir la vista hacia ellos al margen de su función de sintetizadores de intuiciones, igual que puede desmontar una fresadora aunque no esté realizando su función de fresar cosas. Sin embargo, si el argumento de Kant en la «Analítica» sobre la naturaleza de intuiciones y conceptos es correcto, no

podemos hacer nada parecido. Porque la tesis de ese argumento es que un concepto no es una cuasi-sustancia, sino, si se quiere, una actividad. Además, desde el puro sentido común, es difícil encontrar una operación que uno pueda realmente realizar y que cuente como «descomponer un concepto en sus partes». (Si no, inténtenlo.)

Pese a todo, está claro que realizamos algún tipo de operación sobre algún tipo de objeto cuando descubrimos y formulamos proposiciones analíticas. ¿Qué operación es esa, y qué objetos? La respuesta que se ha hecho habitual en la filosofía contemporánea, y que (pienso yo) es la correcta, es que examinamos *palabras*. Una de las operaciones que encierra dicho examen es observar cómo se usan y descubrir qué diríamos y qué no diríamos. En ese proceso descubrimos, por ejemplo, que no diríamos de ningún soltero que está casado, de ningún cubo que es redondo, o de ninguna proposición analítica que es a posteriori. Una segunda operación es acuñar palabras nuevas, o dar sentidos nuevos a las viejas. Lo hacemos de forma que el patrón de uso de esas nuevas palabras, o de las viejas usadas con nuevos sentidos, nos ayude a evitar problemas verbales a que nos habían conducido nuestros usos antiguos. Mediante estos dos tipos de operación, descubrimos qué proposiciones son analíticas según usos pasados, y creamos proposiciones analíticas nuevas al idear usos futuros.

Lo que a mí me gustaría es convencerles de que la caracterización más útil de eso que Kant quiere hacer en la *Crítica* es en términos de dichas operaciones. En la realidad, lo único que puedo hacer es lo que ya vengo haciendo: a saber, señalar con el dedo las dificultades que tiene cualquier otra interpretación. Con todo, déjenme intentar hacer un poco más eligiendo una muestra del tipo de cosa que hace Kant y enseñándoles que interpretar lo que él discute como un ejercicio de análisis lingüístico vale para el caso.

Consideremos las palabras «concepto» e «intuición». Tal como las han usado normalmente los filósofos, y como las usa Kant la mayor parte del tiempo, funcionaban como nombres de cuasi-sustancias. Por ejemplo, tenía sentido hablar de descomponer conceptos en otros conceptos, de los conceptos como herramientas, de las intuiciones como algo que se presenta en multiplicidades como las

uvas se presentan en racimos, etc. De esta forma, la diferencia entre los dos tipos de cosa nombrados se tomaba por análoga, aunque vagamente distinta, a la que hay entre dos tipos de objeto físico: digamos, entre moldes para mantequilla y bloques de mantequilla, o entre prensas de uva y uvas. Así, aceptamos como analíticas proposiciones del estilo de «todo concepto que es diferente a otros es separable de todo otro», «ningún concepto está inseparablemente unido a ningún otro», y similares. Consideremos ahora otra posibilidad: tal vez «concepto» e «intuición» se parezcan más a palabras como «norte» y «sur», o «izquierda» y «derecha». Estos pares de palabras tienen dos características interesantes: (1) no son nombres de tipos de cosa por más que estiremos su significado (su uso primario es como adjetivos); (2) son perfectamente recíprocos, de modo que cualquier ocasión de aplicar uno de ellos será automáticamente una ocasión para usar el otro. Son de naturaleza adjetiva porque no tendríamos ocasión de hablar de norte ni sur, ni de izquierda y derecha, a menos que tengamos algo más de qué hablar aparte de la izquierdez o derechez, o de la nordidad o sudidad. Dicho de otro modo, no las usamos porque haya cosas tales como nortes o sures, o como derechas e izquierdas, sino porque hay objetos espaciales cuyas relaciones espaciales mutuas deseamos discutir. Ciertamente, podemos hablar de «el norte» o «el sur», o de «la izquierda» y «la derecha», convirtiendo así en sustantivos palabras que originalmente eran adjetivos, pero, cuando lo hacemos, siempre somos conscientes de que estamos hablando relativamente. Siempre tenemos en mente algún punto de referencia concreto y estamos siempre dispuestos a admitir que lo que es el norte desde un punto de vista, es el sur desde otro, y que lo que es izquierda desde un punto de vista, desde otro es derecha.

Ahora, supongamos que tratamos las palabras «concepto» e «intuición» como meras formas hipostasiadas de los adjetivos «conceptual» e «intuitivo», y que tratamos a estos como recíprocos del modo en que lo son norte-sur y derecha-izquierda. Entonces, no solo será analíticamente verdadero que «conceptos» e «intuiciones» no son cuasi-sustancias, sino también que son meras abreviaturas de descripciones de casos de conocimiento, descripciones hechas

desde dos perspectivas diferentes. Podemos ver nuestro conocimiento de la ley de la gravedad o del color de este pupitre, bien desde el lado de la unidad, bien desde el de la pluralidad. Cuando lo miramos desde el lado de su unidad, pensamos en él meramente como una disposición de conceptos. Cuando lo miramos desde el lado de su pluralidad, lo vemos referido a un agrupamiento de impresiones sensoriales pasadas, presentes y futuras. Si lo miramos exclusivamente desde el primer lado, nos volvemos racionalistas; si desde el otro exclusivamente, nos convertimos en empiristas. Si lo vemos desde ambos lados, y nos damos cuenta de que los lados desde los que miramos no son ingredientes de lo que estamos viendo, sino simples perspectivas de ello, entonces podemos evitar los dos extremos. Los racionalistas insisten en que todas las intuiciones son solo versiones confusas de conceptos; los empiristas, en que todos los conceptos son solo versiones confusas de intuiciones. Pero esto es tan tonto como sugerir que todas las vistas de una mesa desde el lado izquierdo son meras versiones confusas de sus vistas desde el lado derecho, o que insistir en que todos los movimientos hacia el norte de Filadelfia son meras versiones confusas de movimientos hacia el sur. El tratamiento de Kant del conocimiento como una síntesis de intuiciones por medio de conceptos nos hace ver que es un error usar las palabras «concepto» e «intuición» como si fueran las partes de que está hecho un todo más grande llamado «conocimiento» o «consciencia». Tan falso es que sean partes de casos de conocimiento, como que derechez e izquierdez sean partes de esta mesa, o que nordidad y sudidad sean partes de Filadelfia. Todo caso de conocimiento puede mirarse de cualquiera de las dos formas —como conceptual o como intuitivo—, pero las formas de mirar a algo no son partes de la cosa a la que se mira. Todo caso de conocimiento puede describirse desde cualquiera de los dos puntos de vista —exclusivamente en términos de conceptos o exclusivamente en términos de intuiciones —, pero tales descripciones no lo serán de dos tipos de materiales que entran en la composición del conocimiento, sino de dos funciones que el conocimiento satisface.

Por resumir este punto: lo que digo es que la mejor forma de enunciar una de las principales lecciones que nos transmitió Kant en la *Crítica* es como una tesis sobre las palabras «concepto» e «intuición». Kant no descubrió ningún hecho empírico nuevo acerca de la naturaleza del conocimiento, solo un psicólogo puede hacer eso. Tampoco descubrió nada sobre la naturaleza de su ego trascendental, solo un metafísico podría, y Kant no era ningún metafísico. Lo que hizo fue permitirnos ver que el vocabulario que había brotado entre los filósofos en su tentativa de hacer nuevos descubrimientos psicológicos y/o metafísicos se había vuelto un íncubo. Poco a poco se habían endilgado a sí mismos un vocabulario que no solo *no* les ayudaba a descubrir cosas, sino que les estaba impidiendo ver los puros hechos. Arrancando de ciertos hechos obvios —por ejemplo, nuestro conocimiento de que este pupitre es marrón, de que todo lo que ocurre tiene una causa, de que los hombres son agentes libres, etcétera—, se las arreglaron para construir un conjunto de rompecabezas verbales que, al final, les impedían ver esos hechos patentes. Los filósofos, al hablar de conceptos e intuiciones de la misma forma en que hablan de prensas de uva y uvas, en vez de como hablan de diferentes perspectivas o de diferentes direcciones, se han fabricado sus propios problemas verbales. Se han construido una imagen con dos conjuntos de cuasi-sustancias que existen codo con codo en el espacio mental; a fuerza de trabajar dentro de los límites de esa imagen, se han forzado a sí mismos a dos posiciones insostenibles y paradójicas, a saber, el dogmatismo racionalista y el escepticismo empirista. Por citar de nuevo a Wittgenstein: «una *figura* nos tuvo cautivos. Y no podíamos salir, pues reside en nuestro lenguaje y este parece repetírnosla inexorablemente».[9] La grandeza de Kant estuvo en haber hecho posible que nos libráramos de esa imagen.

He presentado este esbozo de un posible cambio en nuestro uso de las palabras «concepto» e «intuición» como una sugerencia sobre el modo de interpretar el descubrimiento de Kant. Déjenme repetir ahora que no afirmo que Kant pensara en él de esa forma. Al contrario, hasta el fin de sus días no encontró una alternativa mejor

que seguir usando las palabras «conceptos» e «intuición» como nombres de cuasi-sustancias; creo que eso explica por qué lo que dice en la «Analítica» sobre los conceptos y las intuiciones, aunque de entrada es claro, resulta extraordinariamente desconcertante cuando se estudia de cerca. La razón es que Kant mismo no extrajo nunca la moraleja de su propio descubrimiento de que solo podemos tener conocimiento de objetos espaciotemporales. Esa moraleja es, como he dicho, que las entidades mentales, no siendo espaciotemporales, no pueden ser objetos de conocimiento. Al no extraerla, dijo cosas como que «es tan necesario hacer sensibles los conceptos (es decir, añadirles el objeto en la intuición) como hacer inteligibles las intuiciones (es decir, someterlas a conceptos)».[10] Este pasaje, con su evidente insinuación de que los conceptos pueden existir al margen de su papel como sintetizadores y de que las intuiciones pueden existir al margen de su participación en una multiplicidad sintetizada, es una buena muestra de esos que siguen desconcertando a los kantianos desde entonces. En ellos el problema es que Kant continúa hablando de la conciencia al viejo modo cartesiano: como si la mente fuera un tipo peculiar de réplica del mundo físico, una réplica en la que entidades mentales discretas susceptibles de interacción causal existen en una suerte de espacio mental. Lo desconcertante es que esos pasajes aparecen en las obras del mismísimo filósofo que nos mostró de una vez para siempre que la consciencia de un objeto no es, como pensaba la tradición cartesiana, la reduplicación de este mediante la construcción de una réplica suya. La doctrina de Kant de la unidad trascendental de apercepción debería haber abolido definitivamente la idea del yo como un tipo peculiar de fantasma dentro de la máquina que posee o que construye réplicas fantasmales de los objetos físicos a partir de un material mental llamado «conceptos» e «intuiciones». Dada la teoría de la unidad trascendental de apercepción que presenta en la «Deducción Trascendental», tendríamos que haber sido capaces de dejar de usar la palabra «yo» como si fuera el nombre de un tipo especial de sustancia —a saber, la sustancia mental— que entraba en una clase peculiar de

interacción causal con las sustancias espaciotemporales corrientes. Tendríamos que haber podido porque Kant nos da en la «Deducción Trascendental» un modo sustitutorio de pensar en el significado de «yo» o de «sujeto»; ahí abre la posibilidad de pensar en la consciencia de un objeto como (en expresión suya) simplemente la unidad de nuestras representaciones, no algo además de ellas. En otras palabras, nos mostró que el yo humano no era una cosa más añadida a todas las demás cosas de las que somos conscientes, sino que era simplemente la consciencia de esas otras cosas. Una vez que muestra eso —o, por decirlo como prefiero yo, una vez que propone ese abandono de los viejos modos de usar las palabras «yo», «sujeto» y «conciencia» y sugiere otros nuevos—, a Kant no le toca seguir hablando de los conceptos y las intuiciones como si fueran entidades fantasmales que existen en un espacio mental escrutado por un fantasmal yo. Pero, por desgracia, lo hace.

Que lo hiciera fue un desastre, históricamente hablando. Lo fue porque dio pie a que Hegel, y el siglo XIX en general, anduviera por ahí debatiéndose con el problema de la relación entre sujeto y objeto sin reparar en que el problema se había disuelto en 1787, cuando Kant publicó la segunda versión de la «Deducción Trascendental». Pero, aunque fuera desastroso, también era totalmente natural. Quizá sea mucho pedir que un hombre no solo destruya la imagen tradicional de cómo opera el conocimiento humano, sino que además se dé cuenta de qué es lo que ha hecho. Kant se las arregló para destruir la imagen de la mente y del conocer humano alojada en el lenguaje cartesiano, lo cual quiere decir el lenguaje del sentido común. Lo milagroso de su logro es que el lenguaje en que llevó a cabo la destrucción *era el mismo en que estaba alojada esa imagen*. Leyendo la *Crítica*, uno puede ver a Kant batiéndose desesperadamente contra los límites del lenguaje cartesiano de sentido común y apañándose, al precio de una increíble oscuridad, para superar la imagen alojada en él y anticipar una nueva. Pero *solo* la anticipó. Para que el impacto de la obra de Kant se haga visible en su totalidad, creo que uno debe esperar hasta el surgimiento de la filosofía lingüística en el siglo XX. Los filósofos lingüísticos como Wittgenstein no son más inteligentes o

más profundos que Kant, pero cuentan con una ventaja impagable: son conscientes de las trampas verbales que pone el lenguaje. Saben explícitamente (no solo implícitamente, como Kant) que los problemas epistemológicos no se resuelven con descubrimientos sobre la naturaleza de los conceptos, o de la subjetividad, o de la razón, sino que se disuelven con cambios en el uso de las palabras. Kant todavía se veía a sí mismo haciendo descubrimientos sobre el conocimiento al viejo estilo. Por ese motivo, su forma de describir los problemas que trataba resulta inconsistente sin remedio con las soluciones que él mismo ofrece a ellos. Aunque mucho de lo que Wittgenstein dijo después ya está en Kant, hizo falta la autoconciencia metodológica de Wittgenstein para dejarlo dicho claramente.

Si uno lee a Kant como he propuesto, tomará sus enunciados sobre conceptos, intuiciones, juicios, consciencia y demás, no como descubrimientos empíricos ni metafísicos, sino como enunciados analíticos encubiertos. En tanto que analíticos, su verdad depende de un examen del modo en que usamos las palabras en cuestión o de propuestas de usarlas en formas nuevas y más iluminadoras. Creo que entender a Kant de esta manera le permite ver a uno la relevancia de la observación de Wittgenstein que cité antes: el pasaje en el que dice de su propio *Tractatus* que sus proposiciones se reconocerán finalmente como sinsentidos, una vez se hayan comprendido adecuadamente. Una vez comprendidas, dice, pueden desecharse como una escalera que ya no necesitas. Creo que esta observación se puede aplicar a Kant de la forma siguiente: una vez que nos hemos abierto paso por la solución de Kant a los problemas epistemológicos cartesianos enunciados en lenguaje cartesiano, tenemos que abandonar el lenguaje en que esos problemas venían enunciados y en el que fueron disueltos. Dicho de otro modo, tenemos que abandonar el andamiaje todo de la *Crítica de la razón pura*: las distinciones entre concepto e intuición, entre fenómenos y noúmenos, entre Sensibilidad, Entendimiento y Razón, entre multiplicidad y síntesis, y todo el resto. Si mantenemos dicho andamiaje, nos toparemos con los viejos problemas epistemológicos cartesianos, solo que ataviados de una guisa ligeramente distinta.

Porque entonces, en vez de preocuparnos, con Descartes, por la relación entre mente y materia, nos estaremos preocupando por la de fenómenos y noúmenos. En lugar de preocuparnos, como Descartes, de la relación entre imaginación e intelecto, lo haremos por la relación entre Entendimiento y Razón. En vez de preocuparnos, como él, por la relación entre la Voluntad y el Intelecto, nos preocupará la distinción entre Razón Práctica y Teórica. En otras palabras, si mantenemos la terminología de Kant e interpretamos como algo realmente existente las diversas entidades mentales, facultades mentales, etcétera, que él postula, cosas cuya naturaleza necesita ser explorada, en lugar de como una nueva forma de hablar sobre el viejo asunto de siempre, entonces lo mismo habría dado que nos hubiéramos quedado con Descartes. (La terminología de Descartes al menos es un poquito más simple.) Pero si nos libramos de su terminología, seremos libres de interpretar, no que Kant descubrió algo sobre la naturaleza de la conciencia que a Descartes se le escapó, sino que describió la conciencia de un modo diferente a como aquel la había descrito, un modo que nos libera de los problemas de Descartes. Si lo interpretamos así, podemos ver la *Crítica de la razón pura* como yo sugería al principio de esta charla, como una obra puramente negativa, cuyo único valor reside en remplazar una imagen engañosa del conocimiento humano por otra que lo es menos. Pero si cometemos el error de ver algo más que *solo* palabras en las palabras usadas por Kant para dibujar esa imagen menos engañosa, si las tomamos por nombres de cosas reales que requieren investigación filosófica, entonces sencillamente vamos a necesitar otro Kant que haga por el lenguaje kantiano lo que Kant hizo por el cartesiano.

PRINCIPIOS CONSTITUTIVOS Y REGULATIVOS

Quisiera concluir con otra breve sección. En ella me gustaría mostrar que, al tratar de interpretar el logro de Kant ayudándome de doctrinas y métodos de los filósofos lingüísticos contemporáneos, no hago más que seguir una pista que él mismo da. Hasta aquí he

hablado sobre todo de lo que consigue en su discusión de la naturaleza del conocer y de la consciencia dentro de la «Analítica». Querría pasar ahora a la «Dialéctica» y dirigir la atención a una sección que me parece representar el núcleo de lo que Kant tiene que decir sobre la naturaleza y la finalidad de la filosofía. La crítica de Kant a la especulación metafísica es, desde luego, muy rica y compleja, pero creo que lo esencial de su resultado se resume en su discusión de los principios constitutivos versus regulativos en la sección titulada «El Uso Regulativo de los Principios de la Razón Pura».[11] En ella nos dice que los principios regulativos no necesitan respaldo constitutivo. Como el trabajo tradicional de la metafísica ha sido proporcionar ese tipo de respaldo, la metafísica es superflua. Dicho de otro modo, las reglas últimas que nos dicen qué hacer no requieren justificación en términos de descubrimientos sobre la naturaleza metafísica de las cosas. Lo que Kant tiene aquí en mente queda ilustrado en el caso de los principios metodológicos utilizados en la investigación científica. Esos principios nos dicen que debemos desarrollar nuestras investigaciones empíricas *como si* el universo hubiera sido compuesto por una inteligencia suprema. O sea, nunca debemos estar satisfechos hasta haber encontrado una teoría científica que explique los fenómenos de un modo sistemático, unificado y económico. Tras establecer que debemos seguir este principio, plantea la pregunta de si no necesitamos, como fundamento de tal principio metodológico, una garantía metafísica de que siempre podemos hallar una teoría así, una garantía como la que aportaría una prueba de la existencia de un autor del mundo sabio y omnipotente. Naturalmente, su respuesta no es solo que no la necesitamos, sino que no podríamos tenerla si la quisiéramos, pues la idea de una deidad tal es internamente incoherente. Pero el principal punto sobre el que quiero llamar la atención de ustedes es simplemente su afirmación de que no lo necesitamos. Mientras sigamos el principio regulativo en cuestión, que justifiquemos su adopción en términos de una embrollada metafísica o no lo hagamos resulta totalmente irrelevante. Lo que cuenta son las reglas que sigues, no las razones que das para seguirlas. O bien las ciencias

empíricas continuarán siempre encontrando teorías más completas, unificadas y sistemáticas, o bien no lo harán. Lo único que pueden hacer es seguir intentándolo. Al científico no le ayuda en absoluto que los metafísicos le aseguren que el universo está compuesto de un modo ordenado y sistemático.

Esta tesis de que lo que importa es lo que haces, no tu así llamada justificación metafísica para hacerlo, creo que constituye la intuición central de Kant acerca de la naturaleza y la finalidad de la investigación filosófica. Él aplica esta intuición, no solo al caso de las reglas metodológicas que se siguen en la investigación empírica, sino a las leyes morales. El imperativo categórico, para Kant, es un dato que toca fondo absolutamente. Nada podría mostrar que es falso, ni nada que es verdadero; pues solo las descripciones son verdaderas o falsas, las reglas no son ni lo uno ni lo otro. Eso significa que el imperativo categórico no está «justificado» o «basado» en teoría alguna sobre la naturaleza del hombre, del universo o de Dios, ni necesita estarlo. El imperativo es obviamente compatible con todas y cada una de las teorías empíricas sobre esas materias. No obstante, podría parecer (como le pareció a Kant) que está en conflicto con una determinada teoría metafísica, a saber, la del determinismo universal. Pero la respuesta de Kant a esa teoría metafísica no consiste en refutarla, ni en dar una teoría metafísica alternativa. Las teorías metafísicas no pueden refutarse; solo cabe mostrar que sean absurdas. La respuesta apropiada a una teoría metafísica que parece interferir con un determinado dato que toca fondo, como el imperativo categórico, no es una refutación ni, Dios nos libre, una nueva teoría metafísica. Más bien, es sencillamente una demostración de que la teoría *es* absurda. (O sea, una demostración de que la teoría metafísica en cuestión no es relevante para el dato en cuestión ni podría tener nunca relevancia para él.) Todo lo que puede hacer el filósofo moral con una teoría del determinismo universal es señalar que, aunque el determinismo se aplica a todas las cuestiones de hecho empírico, una regla moral es algo completamente distinto de un hecho empírico, y que ningún hecho empírico puede ni respaldar ni debilitar tal regla. En otras palabras, todo lo que puede hacer es lo que Kant de hecho hizo, a

saber, señalar con el dedo la diferencia entre oraciones que contienen el verbo «es» y oraciones que contienen el verbo «debe». Los enunciados-«debe» son reglas; los enunciados-«es» son descripciones. El determinismo puede ser relevante para las descripciones, pero no lo es para las reglas. El determinismo es verdadero del mundo espaciotemporal. Pero los enunciados-debe no versan sobre el mundo espaciotemporal ni sobre ninguna otra clase de mundo. Puesto que son reglas (imperativos, más bien que declaraciones), no versan sobre nada. Que el mundo empírico, o incluso el mundo descrito por un metafísico, es un mundo de regularidades causales es por completo irrelevante para el hecho de que debo actuar del modo prescrito por el imperativo categórico y de que soy libre para hacerlo. Porque la libertad moral no es una propiedad empírica o metafísica de un yo empírico o metafísico. Por consiguiente, ninguna teoría empírica ni metafísica puede demostrarme que soy libre o decirme que no lo soy. Ser libre es simplemente seguir reglas. La libertad, sostiene Kant, no puede comprenderse. Pero eso no es decir que haya algún misterio en torno a ella, es decir meramente que no es el tipo de cosa que necesite ser comprendida. Debido a que la libertad no es una propiedad empírica o metafísica —debido a que podemos describir todo lo que hay en el universo sin mencionarla nunca—, no tenemos que comprender la libertad. De hecho, pedir una explicación o una comprensión de la libertad es mostrar justamente que uno no ha captado cómo se usa apropiadamente la palabra «libre». Si yo entiendo qué es una regla y la sigo, entonces he actuado libremente y he hecho todo lo que hay que hacer, pedirme que dé una descripción de mi actividad de seguir una regla es una tontería. Tanto como pedirme que aísle la propiedad de la izquierdez que contiene este pupitre por la cual está a la izquierda de esta silla.

No obstante, en este punto podría uno objetar que Kant sí parece tener una suerte de teoría del yo nouménico, y de la causalidad nouménica que ese yo ejerce. A veces se dice, en efecto, que Kant necesitaba postular un yo nouménico a fin de darle un fundamento a la moralidad. Pero esta forma de presentar el asunto es totalmente confundente. ¿Qué significaría postular un «yo nouménico»? ¿Qué

clase de cosa estaríamos postulando? ¿Cómo podríamos postular, a la luz del análisis de la función de los conceptos que nos da Kant, algo que no fuera una sustancia espaciotemporal o una propiedad de una tal sustancia? La noción de un yo nouménico, tomada como noción positiva, es un sinsentido. Pero, tomada de la forma que el propio Kant nos dice que hagamos —como una noción puramente negativa—, es de lo más útil. Como noción negativa, se limita a decir que la palabra «yo», cuando se usa en oraciones como «yo debo hacer tal y cual», funciona de modo diferente a cuando se usa en oraciones como «yo estoy haciendo tal y cual» o «yo estaba haciendo tal y cual». En estas últimas, la palabra «yo» se usa como un nombre de mi yo empírico o, en otros términos, de un cierto patrón de sucesos espaciotemporales. En la primera oración —«yo debo hacer tal y cual»— no se usa como nombre en absoluto. No hay nada nombrado por la palabra «yo» en esa oración. Si uno quiere, puede decir que lo nombrado es el yo nouménico, pero decir eso no es más que un modo pintoresco de decir que no se nombra nada. Como dice el propio Kant, «nuestro entendimiento obtiene de esta forma [a través del concepto de noúmeno] una ampliación negativa. […] El pensamiento problemático [que] deja sitio para tales objetos [nouménicos] solo sirve, *como un espacio vacío*, para limitar los principios empíricos».[12] Por tanto, cuando habla de un yo nouménico y de la causalidad nouménica de la voluntad moral, no está suministrando un conjunto de principios metafísicos constitutivos que supuestamente dan fundamento al imperativo categórico. Más bien, está diciendo simplemente: «¡no confundas las descripciones empíricas con las *reglas*!» Este lema es el valor en efectivo de su doctrina de que los principios regulativos no necesitan principios constitutivos que los respalden, y de su insistencia en que se valen por sí solos.

Pues bien, considero que esta máxima de no confundir descripciones empíricas y reglas es el principio metodológico común que vincula a Kant con la filosofía lingüística contemporánea. Para filósofos como Wittgenstein, problemas filosóficos son los que surgen cuando intentamos hablar de algo en unos términos que ahí

son inapropiados. Dicho de otra manera, son problemas verbales, en cuanto que resultan de emplear un cierto modo de hablar en situaciones donde el modo de hablar apropiado es otro. Su solución consiste, por consiguiente, en dar reglas de cómo hablar. Esto significa que el verdadero trabajo del filósofo es descubrir y crear proposiciones analíticas descubriendo y creando usos de palabras. Los problemas verbales requieren soluciones verbales. Con todo, la idea de que los problemas filosóficos pueden despejarse con solo fijar nuevas reglas para hablar tiene algo de extraño. Suponemos instintivamente que esas reglas nuevas deben tener alguna justificación más allá del hecho de que, si las adoptamos, las cuestiones filosóficas que previamente nos preocupaban dejan ya de preocuparnos. Incluso admitiendo que los problemas filosóficos pueden diagnosticarse como maneras de hablar confundidas, quizá nos parezca aún que ha de haber algún criterio para las teorías filosóficas aparte de que despejen confusiones lingüísticas. Ahora bien, lo que responde la filosofía lingüística contemporánea a esta búsqueda instintiva de justificación y de criterios es que ella misma es producto de una confusión entre descripciones empíricas y reglas lingüísticas. Dicho kantianamente, es una confusión entre principios constitutivos y regulativos. Lo que mostró Kant es que todos los principios constitutivos que había o que pudiera haber serían necesariamente competencia de las ciencias empíricas. Si concedemos que la filosofía no es una ciencia empírica, entonces hemos de conceder que los principios que proporciona son puramente regulativos. Si concedemos eso, es difícil ver qué clase de principios regulativos podría ofrecernos a no ser principios que guíen nuestro uso del lenguaje. Por tanto, Kant y la filosofía lingüística coinciden en repudiar la concepción tradicional de la filosofía como una súper-ciencia. En esa concepción, que nace de Platón, la misión de la filosofía es hallar justificaciones y fundamentos para los principios regulativos que usamos, o que nos gustaría usar, en la investigación científica y en las decisiones morales. O sea, la tarea de la filosofía es proveer principios constitutivos que respalden principios regulativos. Pero, después de Kant, es imposible (o debería serlo) que la filosofía se vea a sí misma haciendo tal cosa. El

sentido de la lucha de Kant por hacer de la filosofía una empresa *crítica*, en vez de dogmática o escéptica, era abandonar la concepción tradicional de que la filosofía podía descubrir algo nuevo. A mí me parece que la filosofía lingüística contemporánea puede declararse legítimamente la verdadera heredera de Kant. Porque la filosofía lingüística contemporánea hace explícito lo que Kant descubrió, pero dejó implícito: que, sobre convicciones últimas como el sentido de la causalidad y de la ley moral, no podemos esperar más justificación que la que ahora tenemos.

---

[1] En lo sucesivo, las referencias de Rorty a secciones de la *Crítica de la razón pura* de Immanuel Kant se ponen entre comillas. —Eds.

[2] N. del T.: en el original, el juego es con la polisemia de «*bear*», que, como sustantivo, significa «oso», y, como verbo, «llevar consigo».

[3] Ludwig Wittgenstein, *Philosophical Investigations*, trad. G. E. M. Anscombe. Nueva York, MacMillan, 1968, § 109. [N. del T.: en esta cita y las demás de la misma obra, seguimos la traducción al castellano de A. García Suárez y U. Moulines, *Investigaciones filosóficas*, Barcelona, UNAM / Editorial Crítica, 1988.]

[4] *Ibid*., § 126-127.

[5] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trad. C. K. Ogden. Nueva York, Routledge, 1988 [1922], 6.54. [N. del T.: seguimos la traducción al castellano de J. Muñoz e I. Reguera, Madrid, Alianza Ed., 1987.]

[6] Ver Immanuel Kant, *Critique of Pure Reason*, trad. Norman Kemp Smith. Nueva York, St. Martin's, 1951, A51/B75. —Eds. [N. del T.: en todos los casos, seguimos la traducción al castellano de P. Ribas, *Crítica de la razón pura*, Madrid, Alfaguara, 1978. En esta cita tal como la reproduce Rorty en el original inglés, la primera cláusula dice «los conceptos sin intuiciones», no «los pensamientos sin intuiciones».]

[7] Ver *ibid*., B135. —Eds.

[8] *Ibid*., B11. —Eds.

[9] Wittgenstein, *Investigaciones filosóficas*, § 115. [N. del T.: en la cita, «figura» es el término elegido por los traductores al castellano para el original alemán «*Bild*»; la versión inglesa lo traduce por «*picture*», que es la palabra («imagen») que el propio Rorty acaba de utilizar en la frase precedente y que vuelve a repetir en la siguiente.]

[10] Kant, *Crítica de la razón pura*, A51/B76.

[11] N. del T.: en la traducción de Pedro Ribas, esa sección (en realidad, el apéndice a la dialéctica trascendental) lleva por título «El uso regulador de las ideas de la razón pura». En este caso traducimos el título directamente de la versión inglesa para mantener la terminología que emplea Rorty en sus comentarios.

[12] *Ibid*., B312; A260. [N. del T.: las cursivas son de Rorty.]

CAPÍTULO 4

LA PARADOJA DEL DEFINICIONISMO

Una manera de describir lo que ocurrió en el Renacimiento, cuando los filósofos desertaron de Aristóteles y se hicieron «modernos», es decir que su idea de dónde se ubicaba lo indefinido cambió. En el pensamiento aristotélico, la indefinición que experimentamos es resultado en parte de la indefinición existente *in rerum natura*; después de Descartes, allí donde se encuentre indefinición se considera culpa nuestra, un signo de nuestra ignorancia y confusión. Particularmente el empirismo, desde Hobbes a Russell pasando por Hume, ha dado por sentado que lo que encontramos es nítido y que la borrosidad está solo en lo que hacemos nosotros. La historia del empirismo ha sido en buena medida la de los intentos de trazar la divisoria entre lo que uno halla y lo que uno hace mediante la divisoria entre lo definido y lo indefinido. El empirismo ortodoxo vio como una catástrofe justamente a aquellos filósofos que cuestionaban que ambas divisorias fueran la misma (por ejemplo, Hegel, Bergson, Dewey).

No obstante, cuando uno trata de reflexionar sobre si las dos distinciones coinciden, se percata de que «definido» tiene muchos sentidos y, por tanto, «indefinido» también. A medida que el lenguaje de las controversias filosóficas cambia, lo hace el sentido de esos términos. Hoy, el combate entre el empirismo ortodoxo y la herejía de la indefinición objetiva se libra principalmente en la filosofía de la lógica, y en ella el sentido que se le da a «definido» es: «susceptible de una aplicación incondicional del principio de tercero excluido». Lo que denominaré «definicionismo» es la opinión de que no hay nada a lo que se pueda llamar razonablemente un enunciado y que no sea ni verdadero ni falso. (Como veremos, la evasiva que supone ese «que se pueda llamar razonablemente» no es fácil de eliminar de la posición definicionista.) En qué sentido los que cierran filas con esta tesis son los herederos intelectuales de

filósofos que en el pasado expresaron en términos metafísicos o epistemológicos su debilidad por lo nítido, me parece que está bastante claro. Donald Williams, el exponente más explícito, coherente y pertinaz del definicionismo, hace expresa esa herencia cuando dice:

> Creo que el universo consiste, sin resto alguno, en el despliegue de sucesos en el espacio-tiempo, y que si aceptamos de modo realista el tejido tetradimensional de facticidades yuxtapuestas, podemos prescindir de todas esas nebulosas categorías no factuales que tanto han importunado a nuestra estirpe: lo potencial, lo subsistente y lo influente, lo nouménico, lo numinoso y lo no-natural.[1]

La mayoría de definicionistas tienen más remilgos a la hora de hacer declaraciones metafísicas, pero no hay duda de que es el disgusto por las «nebulosas categorías no factuales» lo que nutre su fervor moral. Incluso cuando se deja de lado la distinción «lo hecho versus lo hallado» por demasiado metafísica o demasiado vaga, se remplaza inmediatamente por la de lo reducible y lo irreducible y se seleccionan las mismas categorías como candidatas obvias a reducción. Para el definicionismo, «no-factual», «reducible» y «resistente a reconstrucción en lenguaje extensional» vienen a ser prácticamente sinónimos.

Últimamente, la resurrección del pragmatismo, unida al tono anti-reduccionista de la filosofía del lenguaje ordinario, vienen dándoles quebraderos de cabeza a los definicionistas. No es de extrañar, pues el hilo de pensamiento común a ambos movimientos es principalmente el contextualismo, y el contextualismo en teoría lógica es el enemigo natural del definicionismo, lo mismo que el contextualismo en epistemología es enemigo natural del empirismo de los datos sensoriales, y el organicismo en metafísica, enemigo natural del atomismo. A los definicionistas les gusta hacer como que sus oponentes tienen una mente nebulosa e imprecisa que proyecta sobre el mundo su propia imprecisión y nebulosidad, exactamente igual que hacían sus predecesores intelectuales con los idealistas. Esto trae como consecuencia que se ataquen las conclusiones más bien que los argumentos, lo que a su vez provoca que la cuestión

del definicionismo se presente como un contraste entre visiones claras y nebulosas de un objeto, y no como un desacuerdo (que eso es en realidad) acerca de cuánto necesita uno saber sobre los objetos circundantes antes de poder ver apropiadamente un objeto dado.

La palabra clave en el enunciado de la tesis definicionista es «incondicionalmente». El contextualismo dice que han de cumplirse ciertas condiciones antes de poder aplicar el tercero excluido a un determinado enunciado. (Si el «indefinicionismo» existiera como algo más que un pelele para que lo vapuleen los definicionistas, seguramente diría que el tercero excluido es inaplicable a determinados enunciados en *todas* las condiciones.) No obstante, los contextualistas difieren entre sí en lo relativo a qué condiciones requiere su aplicación a qué enunciados. Unos dirían que una condición para que verdadero-o-falso sea aplicable a «mañana habrá un combate naval» es que llegue mañana. Otros lo negarían, pero dirían que la verdad-o-falsedad de «todas mis camisas están en la lavandería» requiere la verdad de «tengo camisas». Otros quizá nieguen las dos cosas, pero aun así insisten en que la verdad de «Alcibíades ha alcanzado el uso de razón» es requisito para la de «Alcibíades es un general o es un no-general» y, por consiguiente, para la verdad-o-falsedad de «Alcibíades es un general». Aún otros puede que sostengan que la verdad de «Epiménides es no-cretense» es requisito para la verdad-o-falsedad de la oración «todos los cretenses son mentirosos» cuando la pronuncia Epiménides.

A todos estos presuntos contraejemplos de su tesis, el definicionista replicará con una de las siguientes estrategias. O bien (1) afirmará que el pretendido enunciado no lo es realmente (por lo cual no tiene nada de raro que no le sea aplicable la verdad-o-falsedad), o bien (2) afirmará que habría que reformular el enunciado de modo que incluya como parte propia la aserción de la verdad de la condición en cuestión. El peligro de usar (1) es que el definicionista puede volver trivial su tesis restringiendo sin más el término «enunciado» a cosas que la cumplan. Si lo hace, la discusión simplemente se traslada a la pregunta «¿qué criterio independiente hay de ser-un-enunciado?» Una vez que se suscita

esta pregunta (nada fácil de responder), la tesis definicionista empieza a parecerse más a la propuesta de un lenguaje alternativo que a un enunciado de hecho. De ahí que los definicionistas poseídos por la *Sehnsucht* de una metafísica atomista, de los que nuestra cita de Williams es ilustración, normalmente adopten la segunda estrategia y digan que cada presunto contraejemplo S no es más que una forma confusa y elíptica de un enunciado S' al que se le aplica la solución dada en (2). La táctica restituye el carácter cuasi-fáctico de la tesis definicionista al presentar el contraste entre definicionismo y contextualismo, no como dos propuestas prácticas alternativas sobre maneras eficaces de hablar, sino como la diferencia entre darse cuenta o no darse cuenta de las complejidades del habla común. El contextualista queda retratado como alguien que echa mano de una elipsis casual salida de labios del hombre de la calle y luego le enreda sofísticamente demostrándole que la elipsis viola principios lógicos elementales que suscribe instintivamente.

Ahora bien, es obvio que el punto débil de esta estrategia está en la relación entre S y S', y es ahí donde incide el contraargumento contextualista. En este artículo quiero formular una versión generalizada de ese contraargumento, para mostrar luego cómo se aplica a dos casos representativos de la controversia sobre el definicionismo: las llamadas paradojas pragmáticas y el jaleo levantado por la noción de «presuposición» de Strawson. Terminaré extrayendo la moraleja (puede que obvia, pero a la que ciertamente no se ha prestado atención) de que, mientras se considere que la cuestión es fáctica, el contextualista tiene todas las de ganar, pero que, si el definicionista se aviene a unir fuerzas con el pragmatista y convierte su tesis en un principio regulativo, puede salir victorioso incluso en la derrota[2].

I

Voy a distinguir entre tres réplicas a la afirmación de que cualquier contraejemplo S que viole la tesis definicionista puede remplazarse por un S' que no la viole. La primera es un mal

argumento, la tercera (que establece lo que llamaré «la paradoja del definicionismo») es uno bueno, y la segunda es buena solo en la medida en que prefigura oscuramente la tercera.

La primera réplica es que la posible sustitución de S por S' no se zafa de la existencia de S ni del contraejemplo que proporciona. Que uno pueda construir un lenguaje nuevo, o una fracción de lenguaje, donde S' se da y S no se da es un artificio que nada prueba sobre S. Tampoco la afirmación de que S' expresa la intención de quienes profieren S se puede verificar, salvo apelando a la intuición (cosa que un contextualista resuelto siempre puede frustrar diciendo «eso no es en absoluto lo que yo quería decir; yo quería decir *S*».)

El argumento es malo porque, si se aplicara estrictamente (lo que nunca ocurre), frustraría cualquier explicación de cualquier cosa, tanto en ciencia como en filosofía. A menos que estemos dispuestos a aceptar traducciones y a usarlas, una de dos, o nos condenamos a sucumbir bajo el peso de docenas de vocabularios distintos que para nada ayudan a interpretarse unos a otros, o a poder pensar las cosas únicamente en los términos del primer vocabulario que nos dé por elegir. Lo deficiente de esta réplica es que arguye desde la dificultad de chequear traducciones a la afirmación de que una traducción no puede resolver ningún problema. Un análisis de la explicación que arguyera de esta forma se vería obligado a coincidir con el cardenal Belarmino en que la teoría copernicana no es «realmente» una explicación. Confiar en la apreciación intuitiva (y, por tanto, falible) de los resultados de traducir, antes que insistir en un procedimiento decisorio para el chequeo de traducciones, es un aspecto importante de esa virtud que Peirce denominó «autocontrol lógico».

La segunda réplica que puede dar el contextualista es una forma más sofisticada de la primera. Consiste en plantear la pregunta «¿*S'* significaría algo para quien no hubiera empezado por entender S?» El uso explícito más reciente de este argumento quizá sea el ataque de Strawson a la eliminación de los términos singulares por parte de Quine.[3] La réplica hace más preciso lo que intentaba decir la primera al llamar «artificial», o «ad hoc», a un pretendido S'. Si resulta que

solo podemos usar el lenguaje que incorpora a S' pasando antes por el lenguaje que incorpora a S, es plausible argüir que no se ha remplazado S salvo en el sentido trivial en que un tomista, por ejemplo, podría remplazar la formulación matemática habitual de la física cuántica por una formulada en el vocabulario de la física aristotélica. Para librar de su aparente artificialidad al lenguaje que contiene S' habría que mostrar, o bien que se puede aplicar a esa materia directamente, o bien que puede expresar en su propio vocabulario cómo consigue aplicarse.

La fuerza de esta segunda réplica se verá cuando abordemos la tercera. Tal como se presenta, sin embargo, no es concluyente. Considerado como cuestión empírica, no hay modo alguno de saber si habríamos podido entender el lenguaje-S' si hubiera ocurrido que tropezáramos con él primero. Tampoco es fácil saber, dada la peculiaridad de términos como «entender», qué experimento valdría como confirmación de que podríamos haberlo «entendido». (Análogamente, nunca podemos estar completamente seguros de que un aristotélico inspirado no podría haber desarrollado la física cuántica sin emplear nada del aparato matemático que su desarrollo efectivo requirió.) En el caso de los términos singulares, por ejemplo, un definicionista que quisiera defender a Quine frente a Strawson podría señalar a un ordenador digital como ejemplo de algo cuyo lenguaje no contiene expresiones referenciales y, pese a todo, trata sobradamente bien con la materia que se le asigna. Cabe que el contextualista al que ese ejemplo le parezca un camelo apele entonces al programador del ordenador, quien dirá algo así como «por descontado que piensa, pero es demasiado lerdo para saber en qué está pensando». Puede usar esa respuesta para argüir que el compuesto programador-ordenador es capaz de resolver problemas solo porque el programador *puede* usar expresiones referenciales. Pero no es probable que un debate así llegue a decidir nada.

Lo que puede decidir la cuestión es la tercera réplica. Dice así: «aceptaremos que S' es un sustituto de S si, y solo si, puedes darnos un análisis del enunciado «S' es un sustituto de S» que cumpla con tu propia tesis definicionista». La fuerza de la segunda réplica (y, por ende, indirectamente de la primera) sale aquí a relucir

al pedir al definicionista que sea lo bastante coherente como para aplicar a su propia actividad explicativa los estándares que aplica a la materia de explicación. Si resulta que el definicionista solo es capaz de establecer la afirmación de que «S' expresa lo que intenta S» en un lenguaje que contiene expresiones con el mismo defecto que originalmente se atribuía a S, entonces podemos concluir que solo está barajando otra vez el mismo mazo de naipes, no sacando uno nuevo. Desde la perspectiva de esta réplica, «eliminar» la indefinición es como si te quitas una mancha de la mano derecha frotando con la izquierda, te queda manchada la izquierda, y luego no dejas que la primera sepa qué está haciendo la otra.

Ahora bien, antes de empezar a argumentar que el definicionista no puede de hecho cumplir lo que le pide esta tercera réplica, y que, por consiguiente, está verdaderamente atrapado en una paradoja porque no puede hacer buena su afirmación en ningún caso dado sin contradecirla al mismo tiempo, quiero responder a una objeción que podría hacérsele a la forma misma de un argumento así. ¿Qué más daría (podría uno preguntar) si el definicionista no pudiera explicar su propia explicación en la forma en que desea explicar el explicandum que él ha elegido? ¿Acaso la incapacidad de un investigador para describir su investigación en los mismos términos en que describe su materia de estudio arroja alguna duda sobre sus resultados? En mi opinión, esta objeción no hace más que resaltar la diferencia entre la filosofía y la ciencia. A la ciencia se la juzga en su mayor parte por los resultados, y si consigue que logremos más con menos o que tengamos una teoría del campo unificada, nos da un poco igual si la investigación se hizo con un ciclotrón o con un tablero de ouija. Nos da igual, quiero decir, salvo en la medida en que seamos filósofos y tengamos interés en moldear a la vez método y materia de estudio en algún tipo de unidad estética. Según propondré más adelante, el definicionista puede esquivar todas las objeciones contextualistas convirtiendo su tesis en un imperativo como «ten en cuenta que, para ciertos propósitos, el lenguaje-S' es un instrumento de comunicación más eficaz que el lenguaje-S». Esta es una observación sobre la que no cabrían disputas y que obligaría a definicionistas y contextualistas a

preguntarse exactamente quién ha tenido qué dificultad en comunicar qué cosa. Pero, mientras el definicionista esté dominado por una pulsión estética que le empuja a presentar tesis cuasi-fácticas en lugar de imperativos, queda sujeto al tipo de crítica autorreferencial ejemplificada en la tercera réplica.

II

La paradoja del definicionismo (la tesis que está en el núcleo de lo que hemos llamado «la tercera réplica») es la siguiente: «ningún enunciado que exprese el hecho de que un enunciado S, al que condicionalmente no le es aplicable el tercero excluido, es remplazable por un enunciado S' al que le es aplicable incondicionalmente, será susceptible él mismo de una aplicación incondicional del tercero excluido». Como ocurre con la tesis definicionista original, esta tampoco se puede verificar a priori. Si se pudiera, resultaría ser una tautología cuya fuerza dependería de que se asignaran significados especiales a «condicionalmente» y «remplazable» (del mismo modo que la tesis definicionista se puede convertir en tautología usando un sentido especial de «enunciado»). En realidad, la única forma de hacerla convincente es ir probando con casos representativos de sustitución de S por S'. Así se hará (someramente) en la sección siguiente con dos de esos casos. Pero las tesis de este tipo también se pueden respaldar, o por lo menos clarificar, señalando sus analogías formales con otras más conocidas y más generalmente aceptadas. En esta sección voy a intentar mostrar que resulta fructífero mirar la tesis como la traducción a un idioma particular de diversas tesis filosóficas entrelazadas con las que estamos más familiarizados.

Consideremos aquella inicial reubicación cartesiana de la indefinición —sacándola de la naturaleza e introduciéndola en la mente— a la que me refería en el primer párrafo del artículo. En su forma racionalista, donde mejor se ve es quizá en Spinoza, cuyo monismo le exige eliminar las relaciones externas entre sucesos temporales en favor de relaciones internas entre unas atemporales ideas claras y distintas. Para ello, tiene que postular un dualismo

radical apariencia-realidad y, acto seguido, relacionar la apariencia con la realidad diciendo que «la falsedad consiste en la privación de conocimiento que implican las ideas inadecuadas, o sea, mutiladas y confusas»,[4] y que el tipo de lenguaje y de pensamiento basado en la apariencia es remplazable por otro que «reconstruirá» las «ideas universales» confusas, como la de mal, contingencia o cosa, distinguiendo las «nociones comunes» que (cuando se ven confusamente) sirven para construirlas. Al final, lo que era una captación de la apariencia resultará en captación de la realidad, al modo en que una ilusión perceptual se elimina con un cambio de perspectiva. Pues bien, así como la objeción obvia a la teodicea de Spinoza es que la apariencia del mal es ella misma mala, así también la crítica fundamental a su epistemología es que explicar las relaciones externas como confusión es admitir que la confusión es real y, por consiguiente, que ciertas relaciones externas subsisten, a saber, las que hay entre los modos confuso y no confuso de captar la realidad. Uno no puede decir que la temporalidad es mera apariencia y, aun así, admitir que el proceso de eliminar esa apariencia es temporal él mismo. La razón de que los filósofos nunca se hayan quedado satisfechos con los monismos es que, incluso si todos los demás fenómenos pueden salvarse de un modo que sea coherente con principios monistas, el fenómeno de salvarlos no puede salvarse.

Cuando razonamientos como este hicieron que la distinción apariencia-realidad empezara a sonar absurda, la costumbre racionalista de remitir la indefinición al reino de la apariencia se sustituyó por la empirista de remitirla a lo construido, por oposición a lo dado. Ahora bien, los recientes embates contra el «reduccionismo» han dejado claro que la pregunta «¿esa distinción entre lo dado y lo construido la has construido tú, o te venía dada?» es tan engorrosa como la de «tu distinción entre lo real y lo aparente, ¿es ella misma real o aparente?» Para quien quiera reducir uno de los lados de una distinción al otro y favorito suyo, la pregunta crucial es «¿puedes decirnos, *en términos de tu lado favorito*, cómo se ha conseguido la reducción?» Al hacer esas preguntas, le estamos diciendo al filósofo que no puede arrojar sin

más la escalera que ha utilizado para lograr sus explicaciones porque la escalera es una de las cosas que queremos que se nos expliquen.

Estas conocidas paradojas sobre el componente irracional del racionalismo reduccionista y el carácter no empírico del empirismo reduccionista tienen un evidente parecido de familia con nuestra afirmación de que el definicionismo nunca es capaz en la práctica de deshacerse de la indefinición. Son tan grandes los parecidos que se siente uno tentado a buscar una formulación comprehensiva de los peligros que afrontan quienes erigen dualismos y después quieren derivar un lado de la dualidad del otro. Pero la mayoría de esas formulaciones (por ejemplo, las que salpican las obras de Dewey) son tan sumamente comprehensivas que resultan meramente pictóricas. Lo que necesitamos son nociones metodológicas lo bastante específicas para ayudar a desarticular la discusión de afirmaciones definicionistas concretas, y, aun así, lo bastante neutrales para que su aplicación no se limite a discusiones que se tienen solo en alguna jerga particular. Creo que lo más cercano a esas nociones son las de «regla» e «interpretación» tal como se emplean en algunos argumentos de Wittgenstein en las *Investigaciones filosóficas*.

En esos argumentos (que abarcan más o menos de la sección 68 a la 111, y de la 198 a la 247), Wittgenstein se ocupa de mostrar, entre otras cosas, que si uno se hace el propósito de eliminar los límites borrosos de un concepto o del radio de aplicación de una regla, se expone a quedar atrapado en un regreso infinito. La razón es que el mismo impulso que te llevó a pedir una regla va a hacer que pidas otra regla para interpretar la primera, y así sucesivamente (sección 86). El objetivo de quien quiere eliminar la borrosidad es una regla que de algún modo lleve en sí misma su propia interpretación y, por ende, ponga fin a la serie. Pero las reglas no son así (ni, a fortiori, los conceptos). Como dice Wittgenstein, «toda interpretación pende, juntamente con lo interpretado, en el aire [...]. Las interpretaciones solas no determinan el significado».[5] Wittgenstein concluye que «hay una captación de una regla que *no* es una *interpretación*, sino que se manifiesta, de caso en caso de

aplicación, en lo que llamamos «seguir la regla» y en lo que llamamos «contravenirla».[6]

Estas observaciones sugieren un modo de formular el dilema al que se enfrentan todos los filósofos imbuidos de ese espíritu reduccionista del que el definicionismo es hoy expresión: o bien la reducción será infinitamente larga, al suponer primero una regla que correlacione S con S′, luego otra que especifique el sentido en que debe entenderse «correlación», otra más que interprete esta última, etcétera; o bien ese regreso se cortará en algún punto que permitirá al oponente decir que la borrosidad supuestamente cuestionable se ha cambiado de sitio, no se ha eliminado. Ahora bien, el reduccionista, como ya señalamos (para el caso especial del definicionista), puede salir de esa trampa convirtiéndose en un pragmatista *à la* Dewey (aceptando que toda «transacción» va a involucrar tanto elementos borrosos como no borrosos) y *à la* Peirce (aceptando que la lógica es una disciplina normativa y no descriptiva). Si ahonda en las consecuencias de ese pragmatismo, llegará a la posición que Wittgenstein esboza así:

> Podemos —para una finalidad especial— trazar un límite. ¿Hacemos con ello utilizable ahora el concepto? ¡De ningún modo! Excepto para esta finalidad especial. Tan poco como haría utilizable la medida de longitud «1 paso» quien diese la definición: 1 paso = 75 cm. Y si quieres decir «Pero anteriormente no era una medida de longitud exacta», entonces respondo: perfecto, era una inexacta.—Aunque todavía me debes la definición de exactitud.[7]

Pero, por decirlo otra vez, si el reduccionista está dispuesto a llegar hasta ahí, debe preguntarse quién necesita qué clase de exactitud acerca de qué cosa, y no ofrecer una tesis sobre el lenguaje o la lógica, o incluso un programa completo para construir un lenguaje «más exacto».

III

Una vez que hemos tratado de poner en un contexto más amplio la que decimos ser la paradoja del definicionismo, podemos

descender a casos concretos. El primer ejemplo de debate sobre la pretensión definicionista en que quiero fijarme es la «paradoja de la predicción», un problema que, según autores como Paul Weiss, demuestra que existe un sentido «colectivo» de «o» (es decir, uno en el que «φ (p v q)» no implica «φ (p) v φ (q)») que no puede reducirse a su sentido «distributivo» habitual mientras no «le hagamos algo al campo de ítems conectados por el "o" colectivo»[8], como dice Weiss. La paradoja consiste en nuestra tentación de argüir, a partir del anuncio de que un suceso inesperado ocurrirá dentro de los próximos *n* días, que ese suceso no puede ocurrir el día *n* (pues, en tal caso, no sería inesperado), ni el día *n* –1 (por idéntica razón, una vez que hemos eliminado el día *n*), y, por consiguiente, que no puede ocurrir en absoluto. Evitaremos esa tentación, dice Weiss, si recordamos que las fechas en que puede ocurrir el suceso «aún no están distinguidas en la fecha en que se hace un anuncio previo al último día, y que ya hemos eliminado todos los demás días cuando tratamos el último como un día distinto del resto».[9] Se supone que lo que tenemos que hacer en el campo del «o» colectivo del anuncio, antes de poder hacer la inferencia, es el acto de distinguir esos días.

La posición de Weiss es justo de esas que al definicionista le ponen los pelos de punta. No solo hiere su sensibilidad lógica al impugnar la aplicación incondicional del tercero excluido, sino también su sensibilidad metafísica al atribuir al futuro una indefinición objetiva. Aparentemente, se está reviviendo la afirmación de Aristóteles según la cual que llegue mañana es una condición de la verdad-o-falsedad de «mañana habrá una batalla naval», y Donald Williams nos avisa de que, si permitimos esa afirmación, «nos quedamos sin lógica».[10] «Podemos decir con toda confianza», afirma Williams, «que, por muy corta que se quede "nuestra" lógica respecto de la certeza, cualquier otra cosa se queda mucho más corta, de modo que, hablando en general, ningún argumento puede ser lo bastante convincente como para justificar una violación de "nuestra" lógica».[11] Así, vemos a Quine decir que «esta idea [la de que la "paradoja de la predicción" encierra una

paradoja real] ha llevado incluso al profesor Weiss al extremo desesperado de albergar la fantasía de Aristóteles de que "es verdadero que *p* o *q*" es una condición insuficiente de "es verdadero que *p* o es verdadero que *q*"».[12]

A la vista de estas observaciones, un lector ingenuo (vamos a llamarle Pérez) que se haya aguantado la risa al oír hablar de la «ilogicidad» de Aristóteles y Weiss podría quedarse estupefacto al leer un poco después, en la misma polémica de Williams contra Aristóteles, que este último incurrió en el error garrafal de inferir de «necesariamente ($p \vee \neg p$)» a «necesariamente (p) o necesariamente ($\neg p$)». También podría quedar perplejo si por casualidad tropezara con la compilación de casos de fallo del tercero excluido que da Quine en «Notes on Existence and Necessity».[13] Pérez está molesto con el hecho de que no haya la menor pega en que los enunciados que contienen términos modales, comillas, etc., «violen la lógica», pero sea absurdo que lo hagan enunciados sobre sucesos inesperados y batallas navales futuras. Quiere saber cuál es el criterio de cuándo está bien y cuándo no. Quizá entonces se le explique que ciertas partes del discurso ordinario son «extensionales» —en el sentido de que la lógica se aplica a ellas sin consecuencias contraintuitivas—, mientras que otras partes en las que sí se dan esas consecuencias son «intensionales», pero que los ejemplos de Aristóteles y de Weiss, siendo lisa y llanamente enunciados de hecho, son obviamente extensionales. Pérez (dándose cuenta de que todavía no le han dado un criterio, sino que solo le han reformulado el problema) dice ahora: «ah, entiendo. No estáis discutiendo de verdad si abandonar la lógica o no, sino *dónde* abandonarla». Entonces se le dice que no tenemos que abandonarla en ningún sitio porque podemos reformular todas las partes «intensionales» del lenguaje de manera que se les aplique la lógica sin esfuerzo.

Lo que ha ocurrido en este proceso de educar a Pérez es que se le ha iniciado en el hecho de que, además de las reglas (la «lógica») que le dicen cómo manejar una materia, tiene que conocer también otras reglas que le dicen, o bien (a) cuáles son los límites de esa

materia, o bien (b) cómo meter en cintura las cosas de modo que se puedan incluir en ella. Ahora está listo para ejercitarse un poco en el segundo conjunto de reglas estudiándose cómo disuelve Quine las consecuencias contraintuitivas que Weiss encuentra en la paradoja de la predicción. Si la disolución le parece satisfactoria, ya estará en condiciones de considerar que Weiss, si no es «ilógico», al menos no se ha esforzado lo suficiente para poner el lenguaje bajo el control de la lógica.

Que messieurs les assasins commencent. Quine sintetiza la paradoja en cuestión del modo siguiente:

> La trama, en todas sus plasmaciones, es como sigue: K sabe a partir del momento *t* que se ha decretado que, en un momento *t* + *i* para algún entero *i* menor o igual a un *n* especificado, ocurrirá un suceso único de determinado tipo y que será de su conocimiento, y que se ha decretado además que K no conocerá el valor de *i* hasta (digamos) pasado el momento *t* + *i* − ½. Entonces, K arguye que *i* ≤ *n* − 1, porque, si *i* fuera *n*, K sabría inmediatamente después de *t* + n − 1 que *i* era *n*. Entonces, por el mismo razonamiento con «*n* − 1» para «*n*», arguye que *i* ≤ *n* − 2; y así sucesivamente, concluyendo por último, después de *n* pasos, que *i* ≤ 0 y, por consiguiente, que el suceso no ocurrirá en absoluto.[14]

El error de K, dice Quine, es fruto de confundir

> (i) una hipótesis, hecha por K en *t*, de que el decreto se cumplirá, y (ii) una hipótesis, hecha por K en *t*, de que K no sabrá en *t* + n–1 que el decreto se cumplirá. De hecho, la hipótesis (i), incluso como hipótesis hecha por K en *t*, admite dos sub-casos: la hipotética ignorancia de K y la hipotética conciencia de K del hecho hipotético […]
>
> […] La falacia de K se puede poner más de relieve tomando *n* como 1 y retomando el tema del ahorcamiento. El juez le dice a K el domingo por la tarde que será ahorcado al mediodía siguiente y que no lo sabrá hasta la mañana previa. Sería propio de K quejarse en este punto de que el juez se estaba contradiciendo. Y sería propio del verdugo borrarle a K su petulancia a las 11:55 de la mañana siguiente, demostrando que el juez no había dicho nada más contradictorio que la simple verdad.[15]

Nuestra primera reacción a esto, pienso yo, es que Quine ha salvado al juez de la autocontradicción solo a base de «redefinir a la baja» el término «contradicción» y «redefinir al alza» el de «saber». Nuestra solidaridad con K aumentará si nos imaginamos el siguiente diálogo:

> JUEZ: Decreto que (1) será usted ahorcado a mediodía de mañana, y (2) usted no conocerá este hecho hasta entonces. ¿Quiere pronunciar unas últimas palabras?
>
> K: ¿Qué pasa aquí? Naturalmente que lo sé. Acaba usted de decretarlo.
>
> J: Claro que lo he decretado, pero usted no lo sabe. (Aquí el juez agrega el argumento de Quine.)
>
> K: Está usted retorciendo sin más el significado de «saber». Es verdad que puede llegar un indulto o algo, pero si estira «saber» para que signifique «saber incorregiblemente», entonces la segunda parte de su decreto es trivial. En ese sentido, yo nunca *sabré* que seré ahorcado.
>
> J: Eso es problema suyo. Usted solo podría acusarme de incoherencia si yo hubiera especificado que por «saber» en (2) quería decir «deducible de (1)».
>
> K: Al revés, para evitar la incoherencia debería haber especificado usted que *no* quería decir eso, porque el uso normal de «saber» y «decreto» en contextos como este incluye la presuposición de que el oyente sabe aquello que el hablante decreta.
>
> (En este punto, el diálogo degenera en un debate en torno a si puede hablarse propiamente de una «lógica del uso ordinario»; cuando sintonizamos de nuevo, oímos:)
>
> J: Después de todo, fíjese en qué vendría a parar esa presuposición. Si la hiciera explícita, tendría que ser un enunciado autorreferencial, como «usted sabe que todo lo que digo es verdad», o bien tendría que ser una serie infinita de decretos suplementarios, de forma que, para cada decreto *n*, tendría que haber agregado un decreto *n* + 1 que dijera «usted sabe que *n* es verdad», y así sucesivamente.
>
> K: Bueno, ¿por qué *no* decir algo autorreferencial?

> J: Porque no puedo hacerlo sin decir sinsentidos. Observe que «dice», «decreta» y demás se aplican a enunciados, pero «sabe» se aplica a proposiciones. ¿No esperará usted que confunda niveles de lenguaje solo para salvarle el cuello, no?

Lo que ha pasado aquí es que K entró en la sala de juicios con un conjunto de presuposiciones sobre el uso de la palabra «saber». Él había escuchado de sus compañeros de celda predicciones sobre su destino, expresadas como «tú no lo sabes, pero te van a colgar», y no las había hallado paradójicas. En el contexto de la sala de juicios, sin embargo, esperaba escuchar ordenanzas y no predicciones. Lo que escuchó fue un batiburrillo de ambas que sonaba paradójico. El juez podría darle satisfacción a K haciendo explícito qué presuposiciones debe usar para interpretar qué partes del decreto. Pero, si lo intenta, se endosa a sí mismo una tarea de aúpa.

Simbolicemos como (P) la presuposición de K «escuchar el decreto de un tribunal habilita el tener conocimiento de lo dicho en el decreto». (P) es a la primera parte del decreto del juez lo que una regla de inferencia es a una premisa. El dilema del juez, al final del diálogo, es el de alguien que, o bien debe convertir una regla de inferencia (aquí, la negación de (P)) en una premisa, o bien hacer explícita la distinción entre premisas y reglas de las inferencias. Por expresarlo de otra manera, (P) puede considerarse una regla semántica del lenguaje de K que rige su uso de «saber», y el juez tiene que decirle que cambie esas reglas remplazando (P) por algún (P'). El dilema del juez, por consiguiente, es el de quien no solo tiene que hacer un enunciado metalingüístico, sino explicar también qué *es* un enunciado metalingüístico. No obstante, quizá parezca que su dilema no es gran problema: pues ¿acaso no puede dictarle sin más a K la consabida charla sobre teoría de las distinciones entre reglas de inferencia y premisas, o sobre teoría de los metalenguajes? Esta táctica solo funcionará si K está dispuesto a tragarse unos cuantos «oes» no-distributivos. Porque uno no puede hacer esas distinciones sin utilizar un lenguaje intensional. En concreto, uno no puede referirse al significado de expresiones, ni a reglas sobre su uso, sin emplear comillas o algún sustituto suyo, ni, por tanto, sin

crear contextos en los que la distribuibilidad *salva veritate* fallará. Por consiguiente, si K es lo bastante listo para hacerse el tonto y fingir que solo entiende lenguajes extensionales, puede conducir al juez a un regreso horizontal infinito de enunciados del lenguaje-objeto, cada uno de los cuales pretende dar una interpretación al anterior. Incluso si K está dispuesto a seguir el juego de la noción de metalenguaje, todavía tiene la opción de pedirle al juez una regla semántica con la que interpretar su última observación (que iba expresada en el metalenguaje *n*), llevándole así a hacer una observación en el metalenguaje *n* + 1, etcétera.

No obstante, puede ser que el juez haya oído que los niveles de lenguaje están pasados de moda, y eso a lo mejor le anima a hacer un enunciado autorreferencial del estilo de (P″) «no, usted no sabe que todo lo que digo es verdadero». K contraataca: «vale, entonces ¿cómo sé que P″ es verdadero?» El juez le recuerda (quizá por haber leído lo que dicen Ryle y Lawrence sobre la heterologicidad[16]) que la forma de salir de las paradojas autorreferenciales sin invocar niveles de lenguaje es la siguiente: (a) divide el universo de discurso en dos o más clases; (b) di que el predicado supuestamente autorreferencial puede afirmarse o negarse únicamente de los miembros de una de esas clases; y (c) muestra que está mal aplicado en la supuesta paradoja. Por consiguiente (habiendo leído a Scriven[17]), el juez (a) distingue aquellos de sus enunciados que son ordenanzas de los que son predicciones; (b) explica que la pregunta «¿cómo sé que lo que usted dice es verdadero?» se usa solo cuando se pregunta por predicciones; y (c) clasifica P″ como ordenanza. Esta táctica, como ocurría con la de los niveles de lenguaje, solo funcionará si K acepta «oes» no-distributivos. Porque el juez ha abandonado ahora cualquier tesis con la que los definicionistas pudieran simpatizar. Ha accedido a usar un sentido de «verdadero» en el que los enunciados de la forma «S es o verdadero o falso» son verdaderos solo condicionalmente, donde la condición es la verdad del enunciado «S es un enunciado de tal y tal tipo».

Estos dos conjuntos de tácticas no son la única forma en que el juez podría intentar reconstruir las presuposiciones de K y ofrecerle

otras nuevas. Hay más o menos ($n \times m$) formas de hacerlo, donde n es el número de palabras que aparecen en su decreto bipartito inicial, y m el número de distinciones que pueden invocarse para resolver lo que Carnap llama «la antinomia de la relación-nombre». Prácticamente cualquier distinción valdría —por ejemplo, entre proposiciones y oraciones, entre *nominata* ordinarios y oblicuos, entre entidades de diferente tipo lógico, entre usos de «decreto», o de «saber», o de «verdadero» (o, ya puestos, de «que»)—, aunque probablemente ninguna de ellas le aportará más luz a K que su percepción inicial, a lo largo del diálogo, de que, por alguna absurda razón, el juez estaba usando «saber» de un modo que no quería decir «deducible de lo que he dicho». Nuestras observaciones a las dos distinciones que el juez intenta hacer sugieren, en todo caso, que en algún punto se va a producir un fallo en la distribuibilidad *salva veritate*. Si Pérez (el hombre que no tenía claro cuándo el usar «oes» no-distributivos era estupendo y cuándo no) ha seguido todo esto, lo único que habrá aprendido es que, ante un contraejemplo de la tesis definicionista, siempre puedes hallar el modo de convertirlo en una cuestión de nombres, no de lo que está siendo nombrado o descrito. En particular, si recuperamos el tema de la predicción en el decreto del juez estableciendo $n \geq 2$, podemos decirle a Pérez que hemos remplazado el «o» no-distributivo de Weiss, que conectaba entidades no lingüísticas (días), por otro que conecta expresiones, y su «condición» física para eliminar la indefinición (esperar a que los días pasen) por la «condición» lingüística de asentir a un enunciado que expresa determinada distinción lingüística apropiada. Pero, a menos que Pérez tenga un *parti pris* a favor de atribuir la indefinición a los pensamientos y las palabras en vez de a las cosas, eso puede que no le impresione mucho. Y no debería impresionarle, porque él se comprometió a estudiar «lógica», y ahora está bien claro que elegir dónde ubica uno lo que uno considera definido es algo que, sea lo que sea, no es cuestión de «lógica».

[1] Donald C. Williams, «The Myth of Passage», *The Journal of Philosophy* 48, n.º 15 (1951): p. 458.

[2] N. del T.: como se verá en lo que sigue, el artículo no respeta luego al pie de la letra la estructura de este plan, aunque sí se atiene en lo esencial a sus contenidos. El original mecanografiado presenta incluso alguna inconsistencia en la numeración de los apartados (que los editores han corregido) y una referencia interna errónea sobre el número de ellos (no corregida en la edición en inglés, pero que nosotros hemos rectificado en la traducción). Todo ello confirma que Rorty nunca llegó a pulir del todo el manuscrito con vistas a su publicación.

[3] P. F. Strawson, «Singular Terms, Ontology, and Identity», *Mind* 65, n.º 1 (1956): pp. 433-454.

[4] Spinoza, *Ética*, Parte II, prop. 35. —Eds. [N. del T.: seguimos la traducción de Pedro Lomba: Baruj Spinoza, *Ética demostrada según el orden geométrico*. Madrid, Trotta, 2020.]

[5] Ludwig Wittgenstein, *Philosophical Investigations*, 2.ª ed. Oxford, Blackwell, 1967, § 198. [N. del T.: en esta cita y las demás de la misma obra, seguimos la traducción al castellano de A. García Suárez y U. Moulines, *Investigaciones filosóficas*, Barcelona, UNAM / Editorial Crítica, 1988.]

[6] *Ibid*., § 201.

[7] *Ibid*., § 69.

[8] Paul Weiss, «The Prediction Paradox», *Mind* 61, n.º 242 (1952), p. 269.

[9] *Ibid*.

[10] Donald C. Williams, «The Sea Fight Tomorrow», en *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer*, ed. por Paul Henle, Horace M. Kallen y Susanne K. Langer. Nueva York, Liberal Arts Press, 1951, p. 290.

[11] *Ibid*., p. 285.

[12] W. V. Quine, «On a So-Called Paradox», *Mind* 62, n.º 245 (1953), p. 65.

[13] Willard V. Quine, «Notes on Existence and Necessity», *Journal of Philosophy* 40, n.º 5 (1943), pp. 113-127.

[14] Quine, «On a So-Called Paradox», p. 66.

[15] *Ibid*.

[16] Véase Nathaniel Lawrence, «Heterology and Hierarchy», *Analysis* 10, n.º 4 (1950), pp. 77-84; y Gilbert Ryle, «Heterologicality», *Analysis* 11, n.º 3 (1951), pp. 61-69.

[17] Véase Michael Scriven, «Paradoxical Announcements», *Mind* 60, n.º 239 (1951), pp. 403-407.

CAPÍTULO 5
REDUCCIONISMO

El comienzo de la filosofía moderna se suele fechar en el siglo XVII. Si le pidieran a uno que especificara el rasgo más importante del cambio que sufrieron en aquel momento los supuestos y el método de la filosofía —el cambio que hizo «moderna» a la filosofía moderna—, una respuesta plausible sería que ese siglo fue testigo del triunfo del reduccionismo, y que la filosofía ha estado dominada desde entonces por ideales reduccionistas. Con «reduccionismo» me refiero a la aceptación de las siguientes doctrinas: (1) muchas distinciones que hace el sentido común y que están alojadas en las formas corrientes de hablar son meramente verbales; (2) usar esas distinciones como guía de investigación llevará a pasar por alto ciertos parecidos importantes entre las cosas distinguidas; y (3) el objetivo de la filosofía es penetrar en esos importantes parecidos subyacentes y formular nuevas formas de hablar que los iluminen y pongan al descubierto el carácter meramente verbal de las viejas distinciones verbales. En el siglo XVII, tanto la ciencia como la filosofía se apartaron de una manera de concebir la meta de la investigación y adoptaron otra nueva y radicalmente distinta. La primera era la aristotélica, donde el propósito de investigar era descubrir los principios peculiares apropiados a cada campo de la actividad humana y las esencias peculiares de cada una de un gran número de especies distintas. La nueva concepción fue la de un mundo donde las distinciones de sentido común aristotélicas entre teoría, producción y práctica, y entre esencia y accidente, ya no tenían aplicación. Aquí, todos los rasgos de la experiencia podían hacerse inteligibles por igual y de la misma manera. Se pensaba ahora que el ejercicio del pensamiento, cuando tiene éxito, en el mejor caso desemboca en un único esquema de proposiciones del que cabría deducir todas las apariciones de cada uno de esos rasgos, o, como mínimo, en un único vocabulario de términos descriptivos

en que podrían describirse todos. Se puede indicar en dos palabras hasta qué punto son diferentes esta concepción reduccionista de la investigación y la concepción «cartográfica» aristotélica diciendo que, antes del siglo XVII, si una distinción estaba alojada en nuestras formas corrientes de hablar, eso era prueba prima facie de que resultaba *im*prescindible, mientras que, después, era prueba prima facie de que resultaba *pres*cindible.

Dicho de otro modo, el aristotélico entendía que su tarea había terminado si conseguía formular caracterizaciones generales para cada una de las grandes divisiones entre materias de estudio y entre tipos naturales dentro de ellas. Lo suyo no era subsumirlas en epígrafes más vastos, ni encontrar un único vocabulario en que poder reformular esas distinciones, ni derribar las que había entre los diferentes tipos naturales dando con caracterizaciones más amplias o con términos descriptivos de aplicación más extensa. Al contrario, el mérito intelectual se medía por el número de distinciones que uno podía erigir, no por el de las que podía derribar. Este modo de concebir el propósito de la investigación hallaba cómodo asiento en una confianza espontánea en las distinciones ya disponibles en las formas corrientes de hablar; dada esa confianza, se veía natural que la reflexión tuviera como meta primordial hacerlas explícitas y claras, no eliminarlas o sortearlas. La revuelta reduccionista del siglo XVII lo fue contra esta confianza en las distinciones que proveía el habla común. Su objetivo era reducir el número inconvenientemente grande de teorías regionales, vocabularios y problemas que, según enseñaban Aristóteles y el sentido común, hacían falta para tratar con las diversas regiones en que estaban naturalmente divididos la investigación humana y el mundo. Su esperanza en una tal reducción se veía reforzada por su desprecio del verbalismo, desprecio derivado de la aversión por lo que parecían controversias escolásticas estériles y que se combinaba con un recelo general hacia el antropocentrismo. Ese recelo hizo caer en la cuenta de que el habla común fue creada por el hombre para fines humanos y que, por consiguiente, podía ser que sus distinciones reflejaran intereses humanos nada más, no lo que Spinoza llamó la «naturaleza en sí misma de cualquier cosa», o lo

que Ryle iba a llamar después «la forma del hecho». Ese desprecio por las distinciones meramente verbales, junto con la admiración por las formas no cotidianas de hablar que empleaban los matemáticos y por los extraordinarios éxitos de la ciencia en descubrir teorías cuyo radio de aplicación atravesaba los límites usuales, persuadieron a los filósofos de que la meta de la reflexión filosófica era la síntesis. Se convencieron, además, de que un modo en que podía lograrse la síntesis era desarrollando formas simplificadas de hablar. Estas formas simplificadas sustituirían las múltiples distinciones regionales incrustadas en el lenguaje ordinario por el mínimo posible de ellas. Creían que el vocabulario reducido tendría dos ventajas. Primero, haría imposible formular los problemas concretos que habían debatido los escolásticos y, en general, todos los problemas meramente verbales; en segundo lugar, formular un vocabulario reducido llevaría a reconocer aquellas distinciones últimas que *sí* pertenecían realmente a «la naturaleza de las cosas en sí mismas».

Esta concepción reduccionista de la naturaleza de la indagación filosófica ha sido común a casi todos los movimientos filosóficos modernos (con la posible excepción del hegelianismo). Su manifestación más reciente ha tenido lugar en el contexto de lo que Bergmann llama «el giro lingüístico» en filosofía. También ha sido su manifestación más explícita, pues el giro lingüístico ha puesto al alcance de los filósofos un grado de autoconciencia metodológica que resultaba imposible anteriormente. En particular, el método de filosofar del lenguaje ideal nos ha permitido formular por primera vez criterios relativamente claros para juzgar el éxito o el fracaso de los intentos de síntesis-por-reducción emprendidos por los filósofos. Si te dicen que determinadas entidades metafísicas son reducibles a otras, o que ciertos conceptos son reducibles a otros conceptos, es difícil comprobar la afirmación. Pero cuando los problemas filosóficos se plantean en el modo formal de hablar, la noción vaga de «reducción» puede hacerse bastante precisa.

Un modo de precisarla es el siguiente: podemos definir un *análisis reductivo* como una demostración de la *eliminabilidad de un vocabulario para un propósito dado*. La afirmación de que el vocabulario V —considerado como un conjunto de palabras definido

por enumeración— es eliminable para un propósito dado puede analizarse, a su vez, en términos de *traducibilidad de un enunciado*. Por el momento dejaré sin analizar esta última noción, pero volveré sobre ella más adelante. Aquí diré solamente que considero la equivalencia como la forma más fuerte de traducibilidad, mas no la única. Quiero hacer hueco a otros sentidos más débiles de traducibilidad como, por ejemplo, el que queda cubierto por la relación asimétrica «no dice más que», como en observaciones del tipo: «el opio tiene una virtud dormitiva» no dice más que «el opio normalmente duerme a la gente», o «la muerte es un modo del ser del que la existencia humana se hace cargo tan pronto como es» no dice más que «las decisiones de la gente están afectadas por su conocimiento de que van a morir».

Diré que un vocabulario V es eliminable para un propósito dado si, y solo si, para todo enunciado S usado típicamente para conseguir el propósito en cuestión y en el que se usa un miembro de V, se puede formular un enunciado S' que cumpla las siguientes condiciones:

(1) S' no usa palabras que sean miembros de V (si bien, por supuesto, puede mencionarlas).

(2) S' es traducible a S.

Esta es la definición de «la eliminabilidad de un vocabulario para un propósito dado». Si todo S' al que es traducido un S usa un miembro de un determinado vocabulario V', entonces diré que V es eliminable para el propósito en cuestión *en favor del vocabulario alternativo V'*.

Por tanto, demostrar que un vocabulario es eliminable, ya sea por las buenas o en favor de un vocabulario alternativo, realizando de ese modo un análisis reductivo, es ejecutar dos operaciones en las que puede darse controversia. Primero, debe uno determinar qué enunciados son los que se usan de modo típico para el propósito en cuestión. En segundo lugar, debe mostrar que es posible llevar a cabo las proyectadas traducciones. En consecuencia, los criterios de

éxito de un análisis reductivo serán de dos tipos. Por un lado, habrá criterios para identificar qué propósito cumple un enunciado dado. Por otro, habrá criterios para el éxito de las pretendidas transiciones.

Hay dos puntos en relación con esta definición que merece la pena mencionar. El primero es que está pensada para dejar claro que la unidad de análisis reductivo, a diferencia de la de un análisis no reductivo, es siempre un vocabulario. Ningún reduccionista tendría la impresión de haber reducido nada a nada simplemente por haber traducido un enunciado a otro distinto que no utiliza ninguno de los términos cruciales presentes en el primero. Semejante traducción, si se cumplen determinadas condiciones, podría constituir un análisis del primer enunciado, pero no contribuiría al análisis reductivo si las palabras usadas en S' estuvieran tomadas del mismo vocabulario que las de S. De ahí que a los reduccionistas (como Descartes o Bergmann) les parezca que los análisis que hacen los filósofos no reduccionistas —los aristotélicos, por ejemplo, o los filósofos del lenguaje ordinario— se limitan a explicar lo oscuro por lo igualmente oscuro.

En segundo lugar, algunos análisis reductivos son triviales de manera obvia, como cuando uno muestra que todo el vocabulario del inglés es eliminable para todos los propósitos con que se utiliza, ya que uno siempre puede hablar en ruso. Algunos no son triviales, como cuando muestras que los numerales son prescindibles para el propósito de cumplir las funciones que previamente cumplían los enunciados de la aritmética. En este artículo mencionaré en distintos momentos varios criterios de trivialidad de los análisis reductivos, pero no intentaré formular uno que sea aplicable en todos los casos. Eso no se podría hacer sin dar un criterio general para elegir los miembros del vocabulario cuya eliminabilidad el análisis reductivo pretende poder mostrar. Dicho de otra forma, necesitaría uno un método general para construir el conjunto de compartimentos en que vamos a meter los diversos términos del lenguaje. No puede haber tal método general. Los criterios de lo que va a contar como un vocabulario diferenciado son tan variados como los intereses, las metas y los prejuicios de los filósofos reduccionistas.

Con esta definición en la mano, vuelvo ahora a mi recorrido por los avatares del análisis reductivo en la filosofía moderna. Era natural que el reciente hallazgo de esa claridad que proporcionaba el modo formal de hablar condujera a un auge del filosofar reduccionista. Ese auge fue el movimiento del lenguaje ideal. Pero lo más sorprendente del giro lingüístico en filosofía no es que posibilitara practicar reducciones de modo más eficaz, elegante y seguro, sino que llevara a muchos filósofos a repudiar el reduccionismo en su conjunto y, en consecuencia, a volver otra vez a la concepción cartográfica de la filosofía que el siglo XVII abandonó. Creo que las razones de tal cambio son, a grandes rasgos, las siguientes:

(1) La nueva claridad respecto de las pruebas que debían superar los análisis reductivos de los filósofos solo sirvió para revelar que, por citar a Urmson, «nadie estaba produciendo ningún análisis satisfactorio».[1] En otras palabras, nadie estaba pasando el examen. En todos los casos importantes e interesantes (por ejemplo, en las reducciones fenomenistas), resultó que los enunciados propuestos como traducción sencillamente no cumplían los cometidos que cumplían los enunciados a los que traducían. En el sentido de traducibilidad por el que habían optado los propios reduccionistas —a saber, el de equivalencia—, las traducciones entre vocabularios que fueran distintos de un modo interesante probaron ser inencontrables.

(2) La reflexión sobre (1) llevó a los filósofos a pensarse mejor las metas y los métodos del análisis del lenguaje ideal. Antes, ese modo de filosofar se había tenido por una manera de satisfacer a la vez la necesidad de unificar el lenguaje de la ciencia y de volver científica la filosofía, necesidades ambas que habían estado en el núcleo de la investigación filosófica desde el siglo XVII. Pero ahora, las reducciones practicadas mediante la construcción de lenguajes ideales empezaban a tener pinta de ser solo la última y más taimada argucia que empleaban los filósofos para anunciar descubrimientos cuasi-científicos sin moverse del sillón. Porque empezó a estar claro que el prestigio de los análisis reductivos obedecía al éxito de los análisis basados en descubrimientos empíricos, de los que eran

ejemplo las reducciones dentro de las propias ciencias naturales. Las recomendaciones filosóficas de que era deseable hacer que un vocabulario único remplazara a una gran variedad de ellos empezaron a parecer, en palabras de Ernest Nagel, «predicciones, basadas a menudo en lo que eran conjeturas muy libres, sobre cuál sería la vía de exploración más fecunda en una determinada materia de estudio en un determinado estadio de desarrollo de varias ciencias».[2] La recurrencia de la expresión «en principio» en los análisis reductivos llevados a cabo por los filósofos del lenguaje ideal empezó a sonar sospechosa. A resultas de ello, los filósofos se pusieron en guardia para no confundir la clase de unificación de vocabularios que el descubrimiento empírico hace posible con la que permiten las maniobras verbales, y se preguntaron si esta última podía conducir a algo. Dicho de otro modo, comenzó a parecer que la unificación, no solo de las teorías científicas, sino del lenguaje de la ciencia, quizá habría que dejarla para los científicos. Comenzó a parecer, por ende, que la necesidad de unificar el lenguaje de la ciencia y la de hacer científica la filosofía tal vez tenían que satisfacerse de manera completamente independiente la una de la otra.

(3) La reflexión sobre (1) y (2) dio pie a que se insinuara que, en realidad, no debería sorprendernos que los análisis reductivos practicados por los filósofos no cuadren. Porque… consideremos qué tipos de traducibilidad hay cuya demostración podría dar apoyo a un análisis reductivo no trivial. Decir que ciertos S son traducibles a ciertos S’ equivale a una de las siguientes tres afirmaciones: (i) ciertos S se pueden traducir a ciertos S’ porque ahora sabemos, sobre bases empíricas, que ciertas expresiones referenciales de S refieren a las mismas cosas que ciertas expresiones referenciales de S’ y solo a ellas; a esto se lo puede llamar una afirmación de traducibilidad *empírica*. (ii) Ciertos S se pueden traducir a ciertos S’ porque S es un teorema de un sistema formal cuyos axiomas son traducibles, con ayuda de definiciones estipulativas, a teoremas de otro sistema formal que contiene a S’ como axioma o como teorema; a esto se lo puede denominar una afirmación de traducibilidad

formal. (iii) Aunque no se aporten descubrimientos empíricos ni pruebas de una subsunción entre sistemas formales, la traducibilidad de los enunciados se puede descubrir en virtud de una inspección de las situaciones en que los empleamos; este tercer tipo de afirmación puede llamarse de traducibilidad *analítica*.

Es de presumir que es solo la afirmación del tercer tipo la que el filósofo reduccionista, *qua* filósofo, está haciendo. Está diciendo que puede demostrar la traducibilidad mediante una inspección del uso. Pero, entonces, está haciendo un nuevo tipo de afirmación empírica, afirmación que a menudo es perfectamente razonable en el caso de análisis no reductivos, pero que resulta implausible prima facie en el de los reductivos. Porque, en este segundo caso, está afirmando que, dada la presencia de dos vocabularios distintos dentro de nuestro lenguaje, uno de ellos es superfluo. En otras palabras, está afirmando que nuestro lenguaje contiene dos instrumentos con el mismo cometido. Si todos los enunciados usados para un cierto propósito pueden remplazarse por enunciados expresados en un vocabulario completamente distinto, mientras permanece constante el estado de nuestro conocimiento empírico y lo que sabemos sobre la subsunción entre sistemas formales, parece extrañísimo que el vocabulario superfluo haya sobrevivido tanto tiempo. Lo que parece probabilísimo, dado que ha sobrevivido, es que los análisis reductivos basados en la traducibilidad analítica fracasen.

Esta serie de consideraciones ha llevado a muchos filósofos a concluir que el reduccionismo que el siglo XVII concibió como meta de la investigación se adecúa admirablemente a la ciencia, pero resulta completamente inadecuado para la filosofía. O, por plantearlo de otro modo, que abandonar el análisis reductivo como método filosófico hará posible que la filosofía y la ciencia se desenganchen por fin la una de la otra. Son consideraciones que, como es lógico, han hecho también que esos mismos filósofos sean más comprensivos con el aristotelismo y con los escolásticos de lo que lo ha sido nadie en trescientos años. Tales consideraciones, y las conclusiones resultantes, están expuestas de modo claro y elegante en los últimos capítulos del libro de Urmson *Philosophical Analysis*. En este artículo quiero ocuparme de algunas de ellas tal como él las

presenta. En la segunda sección sostendré que ciertos argumentos suyos no tienen en cuenta un campo de aplicación importante del método del análisis reductivo, a saber, aquellos casos en que se aplica a los vocabularios técnicos de los filósofos. En la tercera parte argüiré que, si bien Urmson tiene toda la razón cuando dice que, si se juzgan por la propia vara de medir reduccionista, los análisis reductivos más importantes fundados en traducciones analíticas sencillamente no funcionan, uno todavía puede formular para ellos un conjunto más laxo de estándares que ayudará a distinguir entre análisis reductivos útiles e inútiles.

Antes de pasar a la segunda sección, empero, me gustaría cerrar esta con alguna observación general más sobre el papel del análisis reductivo en la filosofía moderna. Me he extendido en ello porque creo que es importante darse cuenta de lo muy radical que sería el cambio que supondría para los presupuestos de la investigación filosófica abandonar por completo el análisis reductivo como método. Si quiero extenderme un poco más es para subrayar la importancia del tipo concreto de análisis reductivos que discutiré en la segunda sección, cuando se emplean como arma en las discusiones entre escuelas filosóficas.

Desde luego, no es poca cosa decir, como dice Urmson, que «no existe tal cosa como el análisis reductivo», y que «la antigua doctrina del empirismo británico de que todos los conceptos no simples son compuestos de conceptos simples debe finalmente desaparecer».[3] Pero, si fuera solo cuestión de abandonar el empirismo británico, puede que no hubiera motivo de alarma. Incluso si se tratara solo de abandonar la búsqueda de esas pocas distinciones últimas, simples, elegantes, que pretendidamente bastan para captar «la naturaleza de las cosas en sí mismas», quizá tampoco sería para tanto. Nos contentaríamos con el grandísimo número de distinciones que se alojan en el habla común y podríamos dedicarnos a levantar su abigarrado mapa con paciencia y buena letra. Pero, si abandonamos el análisis reductivo, ¿qué será de ese repertorio de demostraciones nuestras de que muchos de los problemas en que se ejercitaron los filósofos del pasado —y en los

que algunos de nuestros colegas aún se ejercitan— eran meramente verbales? ¿Qué vendrá a sustituir, en particular, a nuestra costumbre de traducir los enunciados que aparecen en los libros de metafísica, bien a falsedades, bien a trivialidades? Al fin y al cabo, la popularidad del análisis reductivo entre los filósofos modernos no obedece menos a su utilidad obvia como arma contra los adversarios que a la que se le presume como instrumento para exhibir la verdadera forma de los hechos. Si ahora lo encontramos inadecuado para esta segunda tarea, de ahí no se sigue que lo sea para la primera, ni, de hecho, que no sea imprescindible para ella. Por ejemplo, si abandonamos nuestros análisis reductivos de esa exigencia del tomista de que pongamos en orden nuestra ontología antes de seguir hablando, ¿qué vamos a decir, entonces, sobre la naturaleza del ser?

Esta serie de preguntas retóricas puede sintetizarse en otra: ¿podemos abandonar el análisis reductivo como método de descubrimiento filosófico y, aun así, conservar los réditos intelectuales que ha devengado su uso como método para decidir qué cuestiones discutir? Hay dos razones, a mi juicio, por las que esta pregunta requiere atención.

La primera es que cuesta trabajo creer que al menos algunos de los intentos de síntesis-por-reducción ambiciosos que han llevado a cabo los filósofos durante los últimos trescientos años no tengan un valor permanente y positivo. La indignación moral que ha levantado entre los filósofos reduccionistas el análisis del lenguaje ordinario procede en buena parte de la sospecha de que se nos está pidiendo que abandonemos las esforzadas victorias que nuestros predecesores le arrancaron al oscurantismo. Por decirlo mal y pronto, estos filósofos (de los que Bergmann es un buen ejemplo) se temen que, si regresamos al punto en donde empezó Aristóteles —a saber, una mera fe en la utilidad de decir lo que normalmente diríamos—, acabaremos donde acabaron los escolásticos. Incluso si reducciones como las que intentaron los cartesianos en metafísica, los fenomenistas en epistemología y los utilitaristas en moral nos chocan hoy por apresuradas y demasiado simplistas, parece obvio

que consiguieron barrer fuera muchas maneras de hablar que debían ser barridas y que no habría que dejar entrar de nuevo.

La segunda razón es que los filósofos contemporáneos que más desconfían del análisis reductivo quieren conservar el derecho, no obstante, a decir cosas como (por citar a Ziff) que «lo que se dice [en las obras de metafísica] probablemente es falso cuando es interesante, y banal cuando es verdadero».[4] Si queremos decir esto, hay que disponer de un procedimiento para traducir las oraciones que aparecen en esas obras a otras expresadas, bien en lenguaje ordinario, bien en alguna jerga alternativa. Sin él, se diría que es imposible reconocer una banalidad o una falsedad cuando se tiene delante. Por decir lo mismo de otra manera, la actitud ante el reduccionismo por parte de los filósofos del lenguaje ideal y de los del lenguaje ordinario es diferente solo en grado. El filósofo del lenguaje ideal, respecto del número total de vocabularios contenidos en el lenguaje, se siente a disgusto con —e intenta prescindir de— todos salvo una selección relativamente pequeña. Por ejemplo, puede que solo se sienta a gusto con el lenguaje de la lógica y el vocabulario empleado para informar de sensaciones. El filósofo del lenguaje ordinario, respecto de esos mismos vocabularios disponibles, se siente a gusto con todos salvo con una selección relativamente pequeña, a saber, las jergas técnicas de los filósofos. Hasta filósofos totalmente identificados con el dicho de que todo enunciado tiene su propia lógica, y que concederían, si no hay más remedio, que los de los libros de metafísica poseen su lógica propia, aún quieren poder argüir que de algunos de ellos, y de su lógica, se puede prescindir sin peligro. El único candidato que se me ocurre al puesto de antirreduccionista filosófico sin fisuras y coherente es Hegel. Solo una fe tan grande como la de Hegel en la necesidad de todo lo que tiene lugar, en el lenguaje y en cualquier otra parte, y en la imposibilidad de separar cumplimiento y génesis, podría admitir que todos los vocabularios que cualquiera haya usado alguna vez son por igual irremplazables, que ninguna voz se pierde totalmente. Si uno carece de esa fe, y por tanto desea desentenderse de algunos

vocabularios, no veo a mano más método para justificar el hacerlo que el análisis reductivo.

I

Empezaré considerando la siguiente sentencia de Urmson, que cito de un capítulo titulado «La imposibilidad del análisis reductivo»: «los filósofos no pueden reducir nada a nada y, por consiguiente, no puede haber éxitos ni fracasos filosóficos en ese terreno».[5] Como deja claro el contexto del que está extraída la cita, Urmson no quiere negar que una clase al menos de reducción es posible, a saber, la del número de vocabularios distintos que empleamos. O sea, no está negando que el análisis reductivo —entendido como la demostración de que un vocabulario es eliminable para un propósito dado— sea posible. Lo que cuestiona más bien es la idea de que mostrar que un determinado conjunto de palabras se puede eliminar —en el sentido de que podemos encontrar otro vocabulario en el que expresar los enunciados previamente expresados en el supuestamente eliminable — sea mostrar nada más allá de eso. Concretamente, está negando que ello demuestre que no existe un determinado tipo de cosa, a saber, el tipo o los tipos de cosas a las que se refieren algunas o todas las expresiones referenciales que son miembros del vocabulario eliminado. Dicho de otra forma, Urmson protesta aquí contra la tendencia a pasar del descubrimiento de que «podemos decir todo lo que queramos sobre A sin usar el sustantivo S» a la conclusión de que «no hay S's». Por citarle de nuevo: «no es tanto que el análisis reductivo tuviera éxito en unos pocos casos y en otros no, cuanto que no hay tal cosa como el análisis reductivo. Que Russell lograra analizar los números no mostró que en realidad no había números, y que no se lograra analizar las naciones como sumas de personas no mostró que había naciones además de personas en un sentido metafísicamente significativo».[6]

Ahora bien, hay un sentido obvio en el que debemos aceptar esta conclusión. No se puede pasar de la reducción en el número de palabras con las que somos capaces de llevarnos bien a la reducción de los contenidos del mundo, si por «reducir los contenidos del

mundo» se entiende el tipo de cosa que uno logra volando por los aires una ciudad o exterminando una especie entera de aves. Los cambios en nuestros modos de hablar no tienen relevancia causal en la aniquilación de entidades no lingüísticas, salvo en el sentido demasiado indirecto de que un cambio en el modo de hablar puede llevar a un cambio en el modo de actuar y, por consiguiente, en el entorno. Sin embargo, sí parece que, en algún sentido de «análisis reductivo», queremos en efecto decir que existe tal cosa, que unas veces ha tenido éxito y otras veces no, y que, cuando lo ha tenido, eso nos ha dicho algo más que el que determinado vocabulario era eliminable. Empecemos por este último punto: cuando nos lanzamos a demostrar que un determinado vocabulario es eliminable, ¿qué nos proponemos?, ¿por qué iba nadie a molestarse en hacer eso? Una respuesta tradicional (y creo que correcta) es: para disolver pseudoproblemas que surgen cuando el vocabulario eliminable se utiliza para ciertos propósitos, y despejar así la senda de la investigación. Urmson, en cambio, no responde exactamente esto. Lo que él dice es que la utilidad de tales demostraciones de eliminabilidad es «aclarar cómo se usa el vocabulario en cuestión».[7] ¿Hay alguna diferencia entre las dos respuestas? Por lo menos hay esta: la respuesta tradicional implica que, en algunos casos, dejamos de usar en ciertas oraciones las palabras del vocabulario eliminable (a grandes rasgos, en las oraciones que sirven para plantear pseudoproblemas y responderlos), en tanto que la de Urmson da a entender que, por supuesto, seguiremos usándolo como ahora, pero ahora tendremos más claro cómo estamos usándolo. En otras palabras, su respuesta sugiere aquella misma imagen de la filosofía, sosteniendo inocentemente un espejo delante del lenguaje, que transmite el dicho de Wittgenstein de que «la filosofía lo deja todo como está». Mas ese dicho solo tiene sentido si interpretamos «todo» como «todo excepto la filosofía», y no cabe duda de que tal interpretación, que viene a significar que el filósofo se declara oficialmente irreal a sí mismo, hace gala de un exceso de modestia. Pero probablemente Urmson no quiere decir eso, pues suponemos que concedería que una de las consecuencias que legítimamente

podrían seguirse de nuestro «tener claro» cómo hemos estado usando ciertas palabras es que decidamos dejar de usarlas de ciertas maneras. No obstante, se diría que en ese caso habría sido mejor distinguir entre dos tipos de demostraciones de la eliminabilidad de vocabularios: uno provocará un cambio en nuestro uso de las palabras del vocabulario en cuestión y el otro no. El primero será de esa clase de demostraciones triviales consistentes en mostrar, por ejemplo, que se puede hacer matemáticas en francés igual que en inglés. El segundo incluirá todos los casos que tienen algún interés filosófico. Parece claro que, si uno no pensara que al menos algunas de las oraciones en las que previamente se habían usado miembros del vocabulario en cuestión dejarían de tener uso —o, cuando menos, dejarían de tener el mismo uso— una vez se mostrara que el vocabulario era eliminable, jamás se molestaría en ponerse a demostrar la eliminabilidad. De hecho, si el proceso de «aclarar» el uso de una palabra no produjera cambio alguno en nuestro uso de ella, bien podría preguntarse uno de qué sirvió la aclaración. Así, el propio Urmson dice que «el análisis de Russell es muy útil; impedirá, por ejemplo, que pensemos en ellos [en los números] como grupos de unidades invisibles».[8] Pero ¿qué es pensar en los números como unidades invisibles? Asumo que no diríamos de nadie que piensa en ellos de esa forma a no ser que alguna vez usara, o estuviera dispuesto a usar, o aceptara que otras personas usen, enunciados donde la palabra «número» —o palabras que designen números concretos, u otras relacionadas— se emplee de formas que son propias para referirse a grupos de unidades invisibles. Por otro lado, de nadie se habría dicho que se le ha impedido pensar en los números como grupos de unidades invisibles a menos que deje de usar, o ya no esté dispuesto a usar o a aceptar el uso, de palabras-número en esos enunciados.

Concluyo que mostrar que un vocabulario es eliminable, o bien es trivial, o bien es algo que cambia nuestros hábitos lingüísticos. Este hecho queda oculto porque, cuando hablamos del «uso» de una palabra y anunciamos que queremos aclararlo, tendemos a excluir de los datos sobre su uso que tomamos en consideración el que la

palabra tiene en las discusiones filosóficas. Normalmente solo tenemos presente su uso *primario*. Creo que la distinción entre el uso primario de una palabra o vocabulario y el secundario es esencial para comprender la eficacia del análisis reductivo en cambiar nuestro uso de los términos. Pero se trata de una distinción que, pese a tener una base intuitiva bastante firme, resulta difícil de precisar mínimamente. A grandes rasgos, puede decirse que el uso primario que se hace de un vocabulario es el uso de sus miembros en enunciados que cumplen con el siguiente criterio: si dejáramos de hacerlos, poco a poco dejaríamos de usar los miembros del vocabulario en cualquier otro enunciado. En los casos en que todos o la mayoría de esos miembros son expresiones referenciales, podemos precisar el criterio un poco más y decir que dichos enunciados son los que nos hablan de los objetos a los que se refieren tales expresiones. En la mayor parte de casos, probablemente no nos equivocaremos si decimos que el uso primario de un vocabulario es el que se pone de manifiesto en aquellos enunciados con miembros suyos que en la práctica hacemos con más frecuencia. Pero este último criterio puede fallarnos por motivos triviales: por ejemplo, quienes no son científicos pueden tomar prestado el vocabulario técnico de una nueva teoría de alguna ciencia con el fin de suministrar metáforas o ejemplos llamativos. Ninguno de estos criterios sirve de mucho para decidir casos fronterizos, pero creo que, para la mayoría de vocabularios, en la práctica no tendríamos dificultad en identificar su uso primario. ¿Con qué propósitos se usan palabras como «bueno», «justo», «espléndido» y «noble»? La verdad, para muchísimos, pero si dejaran de usarse para elogiar nos inclinaríamos a decir, o bien que no tenía ningún sentido usarlas, o que han cambiado completamente de significado. ¿Para qué se usan las palabras «león», «unicornio», «tigre» y «pitón»? Para montones de propósitos, pero si dejaran de usarse para referirse a diversos animales reales y ficticios, nos inclinaríamos a decir lo mismo. ¿Con qué fines se usan «mil», «suma», «multiplicar», «igual a» y «fracción»? Con toda clase de fines, pero si dejaran de usarse para expresar verdades aritméticas, nos inclinaríamos a decir lo mismo.

Ahora podemos puntualizar (como quizá sea obvio) que la mayoría de análisis reductivos son demostraciones de la eliminabilidad de un vocabulario para los propósitos que dicta su uso primario. Cuando hablamos del uso de una palabra y anunciamos que queremos aclararlo, por lo general solo tenemos en mente su uso primario. Cuando mostramos, a la manera de Russell, que el vocabulario de las palabras-número es eliminable, estamos mostrando que lo es para ese uso. Dicho de otra forma, mostramos que es eliminable *para el propósito primordial con el que se utiliza*. No hemos demostrado que lo sea para cualquier otro propósito ni, por tanto, para al menos algunos de sus usos secundarios. Al contrario. Todo el valor filosófico del análisis de Russell es hacernos ver que diversas preguntas filosóficas sobre los números solo pueden formularse y responderse usando oraciones en las que aparecen las palabras del vocabulario supuestamente eliminable, y que no podemos hacerlas cuando estas se traducen en términos del nuevo vocabulario que él propone. Al mostrar que las palabras-número son eliminables en su uso primario, lo que hace Russell es subrayar su ineliminabilidad para algunos de sus usos secundarios, a saber, para hacer dichas preguntas filosóficas y responderlas.[9] Con ello, está sugiriendo que esos usos secundarios son objetables, pues ahora se ven como usos que dependen meramente de utilizar ciertas palabras en lugar de otras. El análisis de Russell, en suma, tiene como efecto decirnos que determinadas preguntas que se formulan y se contestan utilizando oraciones que emplean palabras-número son preguntas *meramente verbales*.[10] Así, el argumento de Russell sobre su carácter meramente verbal es que existe un vocabulario que permite formular los enunciados que realizarán el mismo cometido que cumplían los que ponían de manifiesto el uso primario de las palabras-número, pero que no permite formular los que cumplen el cometido de los que ponen de manifiesto determinados usos secundarios suyos. De este modo, su argumento tiene la misma estructura que uno que concluyera que los problemas sobre el significado cabalístico de ciertas palabras —problemas que se dan en idiomas como el griego o el hebreo, donde las letras del alfabeto se

emplean como numerales— son meramente verbales. Lo son porque no se pueden enunciar en idiomas donde los números se designan por otros medios, idiomas que son tan eficaces o más que el griego o el hebreo para los usos primarios de las palabras-número. Por consiguiente, somos libres de resolver tales cuestiones cabalísticas como gustemos a base de ajustar nuestro uso de las palabras, porque hacemos esos ajustes sin miedo a que nuestra capacidad de hacer matemáticas se vea afectada por ello.

En este punto podemos regresar a la pregunta con la que empezamos: en qué sentido, si es que lo hay, pueden los filósofos reducir el universo. Concediéndole a Urmson que no pueden suprimir piezas del mobiliario del mundo demostrando la eliminabilidad de vocabularios, sin embargo no es casualidad que a menudo expresemos el resultado de tales demostraciones en enunciados de la forma «en realidad no hay X's». Porque un uso primario de los enunciados de esa forma es darnos seguridad de que no tenemos que responder a ciertas preguntas sobre los X's. Así, si vuelvo de Arabia y la gente me incordia con preguntas sobre los hábitos de apareamiento del fénix, el modo normal de librarme de ellas es decir que no existe tal ave. Pero la necesidad de asegurarnos de que no tenemos que responder a ciertas preguntas sobre los X's es, como he estado arguyendo, la principal razón por la que intentamos demostrar la eliminabilidad del vocabulario en el que aparece la palabra «X». Así, si vuelvo de mis indagaciones epistemológicas y alguien me pregunta por lo que he aprendido sobre los poderes del intelecto agente, me libro de esas preguntas diciendo que he descubierto que no hay semejante cosa. Análogamente, si me piden mi parecer sobre la operación de las causas finales, me libro de ello diciendo que he aprendido que dichas causas sencillamente no existen. A diferencia del caso del ave fénix, en ninguno de estos otros estaremos informando de un descubrimiento empírico, pero aun así normalmente usaremos expresiones como «no existen los X's» o «no hay tal cosa como X» para informar del resultado de nuestro análisis. En ambos casos podríamos decir con propiedad, contradiciendo a Urmson, que los filósofos han reducido algo a algo. Porque, en los dos, antes del análisis no había simplemente un modo

de hablar, sino una cierta materia de la que hablar, a saber y respectivamente, el intelecto agente y las causas finales de los fenómenos naturales. Tras los apropiados análisis reductivos, esas ya no son materias para la investigación. Dicho de otro modo, en algunos casos de demostración de la eliminabilidad de vocabularios, nos vemos llevados a concluir no simplemente que podemos hablar sobre el tópico A en el vocabulario V_B igual de bien que en el vocabulario V_A, sino además que no hay ningún tópico A del que hablar, solo el tópico B. Conclusiones como estas son las que, de facto, constituyen la principal contribución que los filósofos han hecho al progreso intelectual, porque son las que cambian las preguntas que se hacen los hombres, no simplemente las respuestas que dan a ellas.

Un modo en que los filósofos han expresado dichas conclusiones ha sido decir que han reducido los A's a B's, o la A-eidad a la B-eidad. En mi opinión, tal forma de expresarse es del todo inobjetable cuando lo que quiere decir quien hace la reducción es que un vocabulario es eliminable para todos los propósitos, porque piensa que todos los usos de las palabras que lo conforman son cuestionables. Llamemos a esos casos *reducciones radicales*. Mientras que los análisis reductivos corrientes pretenden mostrar solo, como he dicho antes, que el empleo de un vocabulario para determinado propósito es cuestionable —por ejemplo, el propósito que ejemplifica su uso secundario entre los filósofos—, una reducción *radical* está pensada para demostrar que es cuestionable en *todos* sus usos. Mostrar que un vocabulario es cuestionable en un uso dado es mostrar, o bien que todas las preguntas y enunciados que manifiestan ese uso son preguntas meramente verbales o respuestas a ellas, o bien que tales preguntas y enunciados pueden traducirse ventajosamente porque existe otro vocabulario en el que pueden plantearse de forma más eficaz. En el caso de la reducción radical, el reduccionista quiere mostrar que todas las oraciones en que aparecen esas palabras se emplean para plantear y responder preguntas meramente verbales o que se pueden traducir ventajosamente. Si semejante afirmación tiene respaldo, no veo por

qué no podría uno resumir los resultados de sus indagaciones reduccionistas diciendo, entre otras cosas, que no hay X's, donde «X» es un sustantivo que aparece en el vocabulario cuya eliminabilidad uno ha demostrado. Desde luego, se podría insinuar que decir eso resulta engañoso porque la expresión «no hay tales y tales» se reserva para casos en que «tales y tales» nombra un género de objetos físicos; pero no creo que en la práctica real esté reservada para eso, y dudo que nadie sufra de hecho ese tipo de engaño.

Con todo, y a despecho de lo que vengo diciendo, la denuncia que hace Urmson de los reduccionistas expresa una queja legítima. Obviamente, hay casos en que comentarios de la forma «no hay X's» son por demás engañosos. A veces el reduccionista no apunta hacia una reducción radical, y sin embargo habla como si lo hiciera. En esos casos, solo cuestiona determinados usos secundarios de la palabra «X», pero no otros, o no sus usos primarios. Ejemplos de ello son «no hay objetos físicos», «no hay mentes», «no hay números», y similares. Son expresiones engañosas porque parecen comprometer al reduccionista con una posición que ni por asomo pretendía sostener: a saber, que todas las preguntas del tipo que normalmente hacemos y contestamos sobre objetos físicos, mentes y números concretos son verbales o es mejor preguntarlas en otros términos. El reduccionista que va por ahí diciendo «no hay números», por ejemplo, normalmente solo quiere dar a entender que ciertas preguntas sensatas que hacemos sobre objetos físicos (como «¿dónde está?») son verbales cuando se hacen sobre números; él está feliz y contento de que otras (como «¿es 17 un número primo?») las sigamos haciendo al viejo modo de siempre.

Los filósofos han creado un montón de problemas permitiéndose usar expresiones de la forma «no hay X's» a modo de elegante figura retórica incluso cuando no querían dar a entender que todas las preguntas sobre X son verbales o es preferible traducirlas. Pero creo que no por eso deberíamos tener la tentación de prohibirlas, o de permitirlas solo cuando queremos anunciar descubrimientos empíricos y no filosóficos. Dicho de otro modo, la chiquillada de caer en exageraciones tontas como «no hay números» no debería

llevarnos a renunciar de antemano a la oportunidad de que algunos análisis reductivos puedan conducir a reducciones radicales y, por ende, a demostrar que no hay X's. A veces sí pueden, si se entiende que la expresión equivale a afirmar que no hay preguntas sobre los X's que merezca la pena hacer. No hace falta que desposeamos a la filosofía de los laureles ganados en los triunfos reduccionistas del siglo XVII solo por escarmentar a unos cuantos materialistas y nominalistas desaforados.

Sin embargo, la tentación de hacerlo es bastante común en nuestros días, y según parece, aunque no está del todo claro, Urmson sucumbe a ella. La tentación consiste en suponer que cualquier afirmación en el sentido de que todas las preguntas que se hacen y se responden usando oraciones que emplean determinada palabra son solo verbales, o traducibles ventajosamente, tiene que ser inevitablemente falsa. (O, en otras palabras, que toda reducción radical debe inevitablemente fracasar.) El argumento habitual para suponer eso es, creo, lo que tiene en mente Urmson cuando dice, en el mismo pasaje que cité antes: «no podemos convertir lo numérico en no-numérico de ningún modo que vaya más allá de eliminar las palabras-número; de lo contrario, ¿cómo podría haber una equivalencia?».[11] Según yo entiendo, este comentario resume la siguiente línea argumentativa: siempre que analizamos el uso primario de un vocabulario dado de forma que se demuestre que este puede eliminarse proporcionando traducciones, estamos concediendo que los términos del vocabulario eliminable sí tienen usos primarios; pero si todos esos términos tienen un uso primario, entonces nada puede demostrar que ese uso no era un uso propio, pues ¿cómo podría mostrarse que era impropio usar un término de la manera en que de hecho se usa primariamente? Concretamente, ¿cómo podría mostrarse que el uso primario de un término estaba en oraciones que se empleaban para plantear y responder preguntas meramente verbales o que es mejor plantear y responder de otras maneras? Concediendo que los análisis reductivos pueden mostrar que algunos usos secundarios de un término son cuestionables, se diría que no puede haber ningún criterio de la impropiedad del uso

secundario excepto su conflicto con el primario o su irrelevancia para él. Así, por ejemplo, el tratamiento de Russell de la aritmética mostró que las preguntas sobre el modo de ser de lo inmaterial eran irrelevantes para el uso primario de las palabras-número. Pero si esto es así, entonces hablar de la impropiedad del uso primario de un término es absurdo y, por ende, también lo es sugerir que todas las preguntas hechas y respondidas mediante el uso de oraciones que emplean ese término son verbales y pueden parafrasearse ventajosamente.

Sin embargo, esta línea argumentativa no tiene en cuenta el caso en que el uso primario de un vocabulario está en oraciones que plantean preguntas. Dicho de otro modo, no tiene en cuenta el caso de una clase de términos a los que denominaré términos temáticos [*subject-matter terms*] y los vocabularios que brotan en torno a ellos. Pienso en la clase de términos cuyo uso primario es delimitar una esfera de investigación. Son los que rellenan el espacio en blanco en contextos como «la teoría de___», «la teoría de Pérez de___», «la naturaleza de___», «métodos para investigar___», «la importancia del estudio de___», y similares. Los triunfos del filosofar reduccionista han consistido en buena parte en eliminar tales términos y sus vocabularios asociados. Esas eliminaciones nos han curado de la costumbre de hacer (lo que ahora llamamos) las preguntas «equivocadas», preguntas que ahora consideramos meramente verbales, pero que un día fueron la médula misma de la investigación. De ese modo, nos han permitido empezar a hacer lo que ahora llamamos las preguntas «correctas».

Considérense los siguientes términos: «el ser», «la contingencia», «la trascendencia», «la materia», «la causalidad final», «el alma», «el mal». Hubo un tiempo en que todos ellos funcionaban como términos temáticos, y hay filósofos —los tomistas, por ejemplo— para quienes aún funcionan así. Sin embargo, gracias a trescientos años de reduccionismo, hay muchos filósofos que se niegan a usarlos de esa manera. Lo que es más importante, se niegan a usarlos en absoluto, aunque puedan mencionarlos, como estoy haciendo yo aquí. Porque su uso como términos temáticos era el primario. Su principal cometido era rellenar el espacio en blanco en

contextos como los que he enumerado más arriba, ayudando a delimitar de ese modo una esfera de investigación. Cuando dejamos de usarlos con ese propósito, dejamos de usarlos por completo.

A quienes todavía los usan les parece una lástima que ya nadie se interese en indagar esas materias. Es tal la perversidad del ego moderno, nos dicen los tomistas, que los verdaderos problemas de la filosofía se han dejado de investigar. Estaría muy bien —prosiguen — que los filósofos analíticos enunciaran explícitamente sus propias teorías, probablemente falsas, sobre qué es el ser, cuál es la naturaleza del alma, etcétera, así podríamos discutir a fondo el asunto; mas su perversidad consiste precisamente en eludir estos temas, pero diciendo cosas que, no obstante, presuponen implícitamente teorías sobre ellos. A esta queja, los filósofos analíticos responden que lo que dicen no puede presuponer tales teorías, ya que no hay nada ahí sobre lo que puedan tener teoría alguna. Replican que son los propios tomistas quienes presuponen teorías implícitamente, probablemente falsas, acerca de otras materias, destacadamente el lenguaje. Pues bien, si «aclararnos sobre cómo usamos las palabras» ha de servir para salir de este impasse dialéctico, el proceso de aclarar no puede consistir en apelar a los usos primarios de los términos en cuestión, o en la asunción a priori de que estos son automáticamente usos propios. Porque utilizar ese criterio llevará siempre a que el debate se decida a favor de los tomistas. Si estoy en lo cierto al decir que, para algunos términos, su uso como términos temáticos es el primario, y si aceptamos la línea argumentativa que he atribuido a Urmson, entonces el filósofo que afirme que todas las preguntas que habitualmente se formulan y se responden con oraciones que usan tales términos, o bien son verbales, o bien se pueden traducir ventajosamente, tiene que estar equivocado siempre.

Para verlo, considérese la frase de Urmson que cité más arriba: «no podemos convertir lo numérico en no-numérico de ningún modo que vaya más allá de eliminar las palabras-número; de lo contrario, ¿cómo podría haber una equivalencia?».[12] Supóngase que remplazamos «numérico» por «ontológico», y «número» por «ser».

Obtenemos, entonces, «no podemos convertir lo ontológico en no-ontológico de ningún modo que vaya más allá de eliminar las palabras-"ser"; de lo contrario, ¿cómo podría haber una equivalencia?» Pues bien, esta última pregunta expresa exactamente el tipo de argumento que de hecho emplean los tomistas para demostrar que, por muy a menudo que cambie el vocabulario de la filosofía, de un modo u otro estaremos siempre dando respuestas, o presuponiéndolas, para las mismas preguntas que hicieron y respondieron Aristóteles y Tomás de Aquino. Dicho de otra manera, por muchos análisis que ofrezcamos de enunciados como «¿qué es el ser?» y «¿es la sustancia la forma primordial del ser?», no podemos escapar de la ontología más de lo que Russell podía escapar de la aritmética. O bien los enunciados generados por tales análisis logran traducir lo analizado, o bien no lo logran. Si lo logran, entonces debe haber un uso primario y, por ende, automáticamente propio, de palabras como «ser», «sustancia» y similares, pues, ¿cómo si no podría haber traducibilidad? El tomista, en otros términos, no aceptará ninguna traducción de los enunciados que componen su temario de problemas ontológicos a menos que esa traducción nos permita formular un temario de problemas ontológicos; ¿cómo podría ser adecuada, nos pregunta, cualquier otra traducción?

Concluyo, por tanto, que, a menos que concedamos que algunos análisis reductivos pueden conseguir mostrar que un vocabulario dado es cuestionable, no ya en algunos de sus usos secundarios, sino también en los primarios, el impasse dialéctico al que me acabo de referir no puede resolverse a favor del filósofo analítico. Si queremos eliminar la ontología demostrando la eliminabilidad del vocabulario del ontólogo, nuestra eliminación no debe proponer meramente un modo nuevo de desempeñar el mismo cometido, sino mostrar que el cometido original no era necesario ni había que intentar cumplirlo. Más en general, cuando demostramos la eliminabilidad de un vocabulario cuyos miembros tienen como uso primario el ser términos temáticos, no estamos proponiendo que haya un modo mejor de hablar sobre el mismo tema, sino que nunca hubo un tema de que hablar. Concluyo, por consiguiente, que si

vamos a dar correcta cuenta de lo que ocurre cuando los filósofos ofrecen análisis reductivos, debemos explicar casos en los que su efecto es un cambio de tema, además de casos en que el efecto sea simplemente ofrecer un nuevo modo de hablar sobre lo mismo. En contra de Urmson, debemos admitir casos donde los filósofos verdaderamente reducen unas cosas a otras cosas, a saber, algún tema a algún otro tema. No obstante, para ello necesitamos una explicación general de los criterios que deciden si una demostración de la eliminabilidad de un vocabulario tiene éxito o no. En concreto, necesitamos distinguir entre varios sentidos de traducibilidad. En la tercera sección de este artículo intentaré esbozar una explicación de ese tipo.

Pero, antes de pasar a ella, me gustaría ocuparme de dos posibles objeciones a lo que he dicho hasta aquí.

Primera objeción: «es engañoso hablar de "el ser" como si fuera *inherentemente* un término temático, y por tanto sugerir que, si deja de usarse para delimitar una materia de investigación, debe por eso dejar de tener uso de ninguna clase. Pues cualquier término puede desempeñar, en alguna fase de la historia de la investigación, el papel de rellenar los espacios en blanco de expresiones como "la teoría de___", o "teorías de___", y/o "la naturaleza de___". Cuando dejamos de hablar de "la teoría de x", el resultado normal no es, en todo caso, que dejamos de usar el término "x", sino que lo trasladamos a contextos como "la teoría x de Y". Así, durante mucho tiempo pudo hablarse con propiedad de "teorías del éter", y esa expresión era equivalente a la expresión alternativa "teorías de la transmisión de la luz". Pero cuando resultó que no hacía falta postular un éter luminífero, empezamos a hablar de la "teoría del éter de la transmisión de la luz". Del mismo modo, se solía hablar de la "naturaleza de las posesiones diabólicas", de "métodos para investigar la posesión diabólica" y similares, pero ahora hablamos de "la posesión diabólica como posible explicación del trastorno mental". Igualmente, antes de la Reforma podíamos hablar con naturalidad de "teorías de la transubstanciación", pero después hablamos de "la teoría de la transubstanciación de la naturaleza de la Eucaristía".

Dicho de otro modo, lo que ocurre con los viejos términos temáticos no es que quedan sin uso, sino que pasan a usarse como nombres de hipótesis —por lo general, aunque no siempre, hipótesis descartadas— sobre el modo apropiado de responder a preguntas sobre un tema nuevo designado por otro vocabulario, a menudo nuevo también. Por consiguiente, es engañoso decir que el análisis reductivo de un vocabulario como el de la ontología tomista, al privarlo de su uso primario, lo priva de todo uso. No necesitamos ver la ontología como un puro error, como la investigación de un tema que, aunque parezca mentira, resultó que no era un tema en absoluto. Más bien, el paralelo con los anteriores ejemplos tomados de la ciencia y la teología nos exige pensar que el vocabulario de la ontología tiene el uso de permitirnos formular una hipótesis (aunque quizá equivocada) sobre un tema describible en algún otro vocabulario. También nos lo exige la cortesía, pues, si lo describimos así, al menos le reconocemos al ontólogo el mérito de ser un hombre que intentaba hacer algo que debía hacerse, pero que a la postre no acertó con las herramientas correctas.»

Así reza la primera objeción. Empezaré concediendo que en las ciencias y, en general, en las disciplinas donde, a diferencia de lo que ocurre en filosofía, hay acuerdo entre los investigadores en cualquier fase dada de la historia de la investigación sobre qué supondría resolver una determinada pregunta, el uso que normalmente se hace de los viejos términos temáticos es convertirlos en términos que rellenen el primer espacio en blanco en expresiones de la forma «la teoría___ de___». Pero creo que este hecho no hace más que subrayar la diferencia entre investigaciones científicas y controversias filosóficas. En las ciencias, los análisis reductivos se desarrollan dentro de un marco de común acuerdo en torno a qué preguntas hay que responder y qué criterios hay para las respuestas; por tanto, el resultado del análisis reductivo no suele ser la eliminación de un vocabulario, sino la coordinación de vocabularios viejos y nuevos dentro del marco de un tercer vocabulario neutral y relativamente permanente que los abarca. Ese tercer vocabulario es, en términos de Carnap, el «marco lingüístico» dentro del cual se resuelven las «preguntas internas» de la ciencia o

ciencias en cuestión; él suministra el vocabulario temático de esa ciencia. Pero la controversia filosófica no tiene este marco lingüístico consensuado dentro del cual trabajar. Pues, a fin de cuentas, ¿sobre qué es hipótesis la ontología tomista sino sobre el ser en tanto que ser?, ¿qué rellenaría el espacio en blanco de «la teoría ontológica de___»? ¿A qué preguntas consideraremos que responde? ¿Es quizá una serie de hipótesis erróneas sobre el comportamiento de las palabras?, ¿o una forma primitiva de filosofía de la ciencia? Tal vez lo sea, pero eso no es más obvio que el que *Lenguaje, verdad y lógica* es en realidad, como algunos han dicho, solo una serie de respuestas a preguntas ontológicas. Desde luego, es posible que una larga secuencia de intercambios dialécticos con el ontólogo nos haga capaces de formular un tercer vocabulario neutral, una caracterización del tema que tanto la ontología tomista como el análisis lingüístico puedan describir como lo que están investigando. Cabe esa posibilidad, pero seguro que no es inminente. En mi opinión, la búsqueda de semejante vocabulario neutral es tan importante y urgente como difícil. Pues si llega el día en que se encuentra, los filósofos estarán un pasito más cerca de esa meta siempre en retroceso que es «convertirse en una ciencia». Pero, hasta que se encuentre el vocabulario neutral, es un error discutir el análisis reductivo sin tener en cuenta los problemas peculiares que presenta el no dar con él. Describir el proceso de cambiar las preguntas que hacemos es, con toda seguridad, mucho más enrevesado que describir el proceso de cambiar las respuestas a esas preguntas. Por eso la metafilosofía está en un estado bastante primitivo en comparación con la filosofía de la ciencia. Pero ambos procesos se dan, y son totalmente diferentes. No tiene sentido intentar fingir que el primero es solo un caso especial del segundo.

Segunda objeción: «las objeciones que usted pone a la opinión de que todo uso primario debe ser un uso propio (y, por consiguiente, de que la filosofía no se puede atrever a decir que el cometido que actualmente cumple un vocabulario no valga la pena) se basan en meter en el mismo saco usos ordinarios y usos técnicos del lenguaje. A nadie se le ocurriría defender esa opinión en relación

con vocabularios técnicos, en especial los de los filósofos. El principio que usted le atribuye a Urmson —que el cometido que cumple un vocabulario vale automáticamente la pena— rige solo, si acaso, para los usos ordinarios de vocabularios no técnicos. Naturalmente, el criterio de si un vocabulario técnico, o un uso técnico de un vocabulario ordinario, desempeña o no una función útil no es si se lo usa o no, sino si cumple la función para la que se inventó. Una palabra como "ser", en el sentido en que la usan los filósofos, es una "palabra de filósofos", y una palabra de filósofos tiene un uso propio solo en caso de que nos ayude a realizar tareas filosóficas.»

Así reza la segunda objeción. Mi réplica se apoya en el mismo hecho que la anterior: el de que no estamos en situación de especificar qué tarea es esa que el vocabulario temático debe realizar. Eso no lo podemos hacer a menos que dispongamos de un vocabulario temático alternativo que nos permita ver los enunciados que exhiben el uso primario del primero como expresiones de una teoría sobre alguna materia. Pero esto es precisamente lo que no tenemos en el caso de las controversias entre movimientos filosóficos. Si pudiéramos formular cuáles son los problemas para cuya solución estaban pensados los vocabularios filosóficos técnicos, entonces sí podríamos tomar una decisión, con bases pragmáticas relativamente claras, sobre si seguir usando o no cualquier vocabulario filosófico dado. Pero, repito, es un rasgo distintivo de las controversias filosóficas el que encierren una discrepancia sobre cuál es el tema que se está discutiendo y, por consiguiente, sobre qué problemas hay que discutir. En estas condiciones, cada parroquia filosófica (léase los tomistas, los analistas lingüísticos, los marxistas, los heideggerianos) posee un vocabulario temático que les permite dar traducciones de los enunciados que hacen cada uno de sus rivales y, por tanto, llevar a cabo análisis reductivos de sus vocabularios técnicos. Cada una de esas traducciones está pensada para mostrar que el vocabulario temático que usan los rivales no delimita en realidad un tema y, por ende, que los problemas que discuten son en realidad meros pseudoproblemas. Cada uno de los rivales así atacados replica que lo inadecuado de la traducción ofrecida queda de manifiesto ya solo por el hecho de que se hace

desde enunciados que se usan para formular preguntas sobre una determinada materia a enunciados que no pueden usarse para ese propósito. En una situación así, los criterios pragmáticos no sirven de nada. En contra de lo que propone la objeción, no podemos indagar por las buenas si la «palabra de filósofos» en cuestión cumple la tarea para la que se pensó. Para ello, deberíamos tener clara la naturaleza de esa tarea. No veo que dispongamos de más método que el del análisis reductivo para discutir la cuestión de si tales tareas son o no meras pseudo-tareas. En la siguiente sección trataré de sugerir que, para discutir dichas cuestiones, lo que se usa y lo que se debe usar es el análisis reductivo.

II

En esta tercera sección del artículo —la última, la más conjetural y (gracias a dios) la más corta— intentaré satisfacer un primer pago de la deuda que contraje al optar por el empleo del concepto no analizado de *traducibilidad* en las dos secciones previas. Recordarán que dije que utilizaría el término «traducible» para designar una relación genérica que mantienen determinados enunciados, una de cuyas variedades es la equivalencia y otra la relación no simétrica de «no decir más que». Ahora necesito explicar esta última y así mostrar que existen variedades de la traducibilidad, más débiles que la equivalencia, que se pueden utilizar para hacer análisis reductivos. Necesito mostrarlo porque tengo que sortear el argumento de Urmson de que no debemos esperar análisis reductivos interesantes (ni, a fortiori, reducciones radicales) a falta de nuevos descubrimientos empíricos, lógicos o matemáticos. Ese argumento ya lo he esbozado en la sección I. Muy a grandes rasgos, dice que, si los vocabularios en que se formulan S y S’ son lo bastante diferentes como para que demostrar que el vocabulario de S es eliminable en favor del de S’ para un propósito dado sea filosóficamente interesante, entonces S y S’ no serán equivalentes. A la inversa, si son equivalentes, entonces los vocabularios en que se enuncian no serán diferentes de un modo interesante. Las premisas cruciales de este argumento son suponer que el lenguaje normalmente no

contiene vocabularios superfluos y, como corolario de ello, que cualquier vocabulario que esté razonablemente bien incrustado en nuestro lenguaje posiblemente está sirviendo para poder formar enunciados que no son equivalentes a ningún otro formulable en el lenguaje. Pienso que este argumento es correcto. Por tanto, para sortearlo tengo que mostrar que pueden lograrse —o, por lo menos, intentarse— análisis reductivos interesantes empleando un sentido de «traducibilidad» más débil que el de «equivalencia».

Creo que el camino más corto será abordar un ejemplo que Urmson utiliza para dar plausibilidad a los supuestos sobre la naturaleza del lenguaje que acabo de mencionar. Es el de la eliminabilidad del vocabulario que usamos para hablar de naciones en favor del que usamos para hablar de personas. Como parece claro que, en algún sentido, en el comportamiento de una nación no hay nada más que el comportamiento de montones de individuos, parece claro que, si el análisis reductivo ha de funcionar en alguna parte, tiene que ser aquí. Pero, señala Urmson, no hay ningún enunciado sobre individuos que haga la tarea que hace «Inglaterra le declaró la guerra a Alemania en 1939». Es así, más o menos, porque el número de acciones de personas individuales que habrían podido constituir tal declaración es indefinidamente grande, y, por tanto, la única oración posible sobre individuos que sería equivalente a la oración en cuestión sería una conjunción indefinidamente larga. Urmson sugiere que en todos los casos similares se dará la misma situación. Los casos similares son aquellos en que tenemos un vocabulario, V′, en el que podemos informar de materias de un cierto tipo K′, y donde también tenemos otro vocabulario V en el que podemos, como si dijéramos, resumir tales informes sobre las materias K′ de un modo que no dice nada definido sobre materias K′, pero sí dice algo definido sobre materias de un tipo K distinto y «superior». Uno de esos casos similares es la relación entre un vocabulario de datos sensoriales y otro de objetos físicos. Tanto en este caso como en el de las naciones, el hecho de que nuestra base para los enunciados sobre materias K siempre sea, o al menos siempre incluya, enunciados sobre materias K′, sugiere que podríamos evitar del todo hablar de materias K y hablar únicamente

de materias K'. Sugiere, en otras palabras, que el vocabulario V es eliminable en favor de V'. Pero el argumento de Urmson es que, puesto que el vocabulario «superior» V no está pensado para decir nada definido sobre materias K', y si lo estuviera sería innecesario y superfluo, casi nunca será posible encontrar una equivalencia entre un enunciado que use términos V y un enunciado de longitud finita que use únicamente términos V'. Dicho de otra forma, el reduccionista está buscando equivalencias entre enunciados de unos vocabularios cuya entera función consiste en no permitir dichas equivalencias.

Esto es una versión muy abreviada y tosca de lo que dice Urmson, pero creo que muestra la fuerza de su posición; como he dicho, no quiero combatir su conclusión de que no pueden encontrarse equivalencias en los casos que él discute. Al contrario, pienso que tiene toda la razón. Lo que quiero preguntar es si existe algún sentido interesante en el que sea verdad decir que «podríamos evitar del todo hablar de materias K y, aun así, no perdernos nada», incluso después de haber concedido que ningún enunciado que use solo miembros de V' es equivalente a uno que use miembros de V. Yo creo que existe ese sentido y que se puede expresar del modo siguiente. Diré que un enunciado S es evidencialmente traducible a otro enunciado S' para un usuario dado si, y solo si:

> Para todo enunciado X tal que ese usuario acepta X como evidencia adecuada para S, y para todo enunciado Y tal que ese usuario acepta S como evidencia adecuada para Y, uno puede construir enunciados S', X' e Y' tales que:
>
> (1) el usuario aceptaría S' como evidencia adecuada para S, X' como evidencia adecuada para X, e Y' como evidencia adecuada para Y;
>
> (2) el usuario aceptaría X' como evidencia adecuada para S'; y
>
> (3) el usuario aceptaría S' como evidencia adecuada para Y'.

Digamos ahora que un enunciado S es *evidencialmente traducible* si, y solo si, para todo usuario del lenguaje existe algún enunciado S'

(no necesariamente el mismo S’ para cada usuario) al que S es evidencialmente traducible para ese usuario. Supongamos que ninguno de los diversos enunciados S’ a los que S es evidencialmente traducible para los diversos usuarios contiene términos que sean miembros del vocabulario V. Entonces, podemos decir que el vocabulario V es eliminable para el propósito de suministrar evidencia sobre materias K’ si, y solo si, todos los enunciados que usan miembros de V y se utilizan para ese propósito son evidencialmente traducibles. Dicho de otra manera, demostrar la eliminabilidad del vocabulario V ya no depende de descubrir equivalencias, sino de hallar pautas de conexión evidencial entre enunciados.

Déjenme que intente ahora desplegar brevemente los conceptos que he embutido en esta definición. Mi definición de la traducibilidad evidencial está pensada para ajustarse a la convicción intuitiva del reduccionista de que nuestra costumbre de clasificar el comportamiento de personas individuales como nuestro vocabulario-de-naciones nos permite hacerlo es una cuestión de conveniencia, pero solo de conveniencia. Desde ese punto de vista, hablar de naciones no es sino un artilugio cómodo para volver más manejables nuestras inferencias desde trozos de comportamiento individual a otros trozos de comportamiento individual. Siendo así, podríamos quitar de nuestro lenguaje los términos-de-nación sin más coste que una pérdida de eficiencia, y no perder por ello ningún conocimiento que hubiéramos podido ganar no haciéndolo. Esta idea queda reproducida en la definición de traducibilidad evidencial porque asegura que, si esta última se da, cualquier cadena de relaciones evidenciales que lleve de un enunciado sobre naciones a otro sobre naciones tendrá su reflejo en una cadena de relaciones evidenciales que lleve de un enunciado sobre personas individuales a otro del mismo tipo. Dicho de otra forma, asegura que, si todos los enunciados que usan palabras-de-nación, y que se emplean para el propósito de predecir el comportamiento de individuos, son evidencialmente traducibles a enunciados que no las usan, entonces no se perderá ningún conocimiento sobre el comportamiento de personas individuales solo porque ya no empleemos palabras-de-

nación. Pues, partiendo de algún S' que sirva como evidencia adecuada de algún S, podremos avanzar por una cadena evidencial en el nivel del habla sobre individuos hasta algún enunciado Y', enunciado que, a su vez, habría sido evidencia adecuada de algún enunciado Y al que hubiéramos llegado partiendo de S y avanzando por una cadena evidencial de enunciados sobre naciones. Por consiguiente, el único coste de atenernos a los enunciados con signo prima y renunciar a los S's e Y's originales sería una menor eficiencia. No obstante, esa pérdida de eficiencia sería muy grande, como es obvio, ya que los S's e Y's con signo prima normalmente serán conjunciones muy largas, si bien ciertamente de longitud finita.

Vuelvo ahora brevemente a la cuestión de si, en el sentido de eliminabilidad que nos da este concepto de traducibilidad evidencial, nuestro vocabulario de naciones es eliminable para el propósito de predecir comportamientos de individuos. Aquí solo puedo decir que no lo sé. No acierto a ver ninguna razón en particular por la que una demostración de eliminabilidad por parte de los reduccionistas no pasara el corte en este caso, pero no sé si lo pasaría o no. La pregunta de si lo haría no es relevante para este artículo, pero mencionarla lleva a otra que sí lo es. La noción débil de traducibilidad evidencial que he desarrollado no debe serlo tanto como para permitir que todas las afirmaciones reduccionistas triunfen automáticamente. En otras palabras, no debe ser el caso que todo vocabulario V sea eliminable para el propósito de predecir el comportamiento de materias K'. A primera vista, podría parecer natural que todos los enunciados que usan términos V y se emplean para predecir el comportamiento de entidades-K' fueran evidencialmente traducibles. Sin embargo, sospecho que esa apariencia engaña, aunque no puedo probarlo. Lo sospecho principalmente porque sospecho que, incluso utilizando esta noción débil de traducibilidad evidencial en vez de la de equivalencia, uno no sería capaz de realizar una reducción exitosa de enunciados sobre objetos físicos a enunciados sobre datos sensoriales. Es decir, dudo incluso de la afirmación más bien restringida, y bastante plausible, de que los enunciados que usan términos de objeto-físico empleados

con el propósito de predecir el comportamiento de datos sensoriales son evidencialmente traducibles. En otras palabras, dudo que sea posible establecer cadenas evidenciales que conecten enunciados de datos sensoriales sin la mediación de enunciados de objeto-físico. Pero, como he dicho, no soy capaz de darles un argumento satisfactorio sobre este punto.

* * *

Esto es todo lo que quiero decir sobre la traducibilidad evidencial. Creo que definir ese sentido de «traducibilidad» nos ayuda a formular criterios de éxito para un gran número de afirmaciones reduccionistas: dicho rápidamente, todas aquellas en las que se afirma que un vocabulario dado es eliminable para el propósito de predecir el comportamiento de un cierto tipo de entidad. El criterio será simplemente si pueden proporcionarse las traducciones evidenciales relevantes. Utilizando dicho criterio, uno puede evitar descartar por triviales o fallidos todos los análisis reductivos no basados en hallazgos empíricos o lógicos, mientras que no hay más remedio que descartarlos si se insiste en que tales análisis suponen descubrir equivalencias.

Sin embargo, este concepto de traducibilidad evidencial no es en sí mismo suficiente para ayudarnos a desarrollar criterios de éxito para los análisis reductivos de vocabularios *temáticos*. Más concretamente, no proporciona un sentido de «traducibilidad» que se pueda emplear en reducciones radicales de vocabularios filosóficos técnicos. Esto es así por dos motivos. Para empezar, el uso primario de los vocabularios temáticos no es predecir el comportamiento de un tipo de entidad. Es más bien formular preguntas y teorías sobre lo que se pretende que es un determinado tema. En segundo lugar, la traducibilidad evidencial, como la equivalencia, es una relación que solo puede regir entre enunciados; no rige entre preguntas. Por tanto, para tener un análisis reductivo plausible de un vocabulario temático, hemos de tener un significado de «traducibilidad» que se aplique a preguntas. Tenemos que poder decir que la pregunta Q no inquiere nada más que otra pregunta Q’,

exactamente igual que tenemos que poder decir que un determinado enunciado no dice más que otro enunciado.

Estas dos consideraciones aclaran por qué los análisis reductivos de vocabularios temáticos resultan mucho más intrincados que los que intervienen en la reducción de naciones a individuos o de objetos físicos a datos sensoriales, que son relativamente directos. En estos casos más directos, resulta relativamente sencillo acordar los diversos propósitos a los que sirven los diversos tipos de enunciados en que se usan los miembros del vocabulario. Por ejemplo, el reduccionista y el antirreduccionista pueden estar de acuerdo en que uno de los propósitos de los enunciados-de-nación es formular evidencia para enunciados sobre la gente. Incluso pueden coincidir en si ese propósito es primario, o lo que es lo mismo, en que si se dejaran de usar para eso, al final dejarían de usarse también todos los demás enunciados que emplean palabras-de-nación. Ahora bien, en el caso de los vocabularios temáticos, el reduccionista y el antirreduccionista sí coinciden en que el propósito primario del vocabulario en cuestión es plantear preguntas sobre un cierto tema. Pero el reduccionista desea decir que no existe tal tema y, por consiguiente, no puede decir que los S-prima que ofrece como traducción de los S's originales servirán al mismo propósito que estos últimos. Supongamos, por ejemplo, que quiero deshacerme de la pregunta «¿qué es el ser?», así como de los enunciados en jerga que los tomistas utilizan para responderla. Obviamente, no puedo afirmar que los enunciados a los que yo los traduzco seguirán contestando a la pregunta «¿qué es el ser?» Pero si demuestro la eliminabilidad del vocabulario del tomista para cualquier otro propósito que no sea su uso primario de formular y responder justamente esas preguntas, el tomista seguirá impertérrito. Por tanto, para que mi análisis reductivo tenga alguna fuerza, tengo que encontrar la manera de argüir, no que yo puedo hacer el mismo trabajo sin usar determinados términos, como en los casos directos de análisis reductivo, sino que no hay ningún trabajo que hacer. Dicho de otro modo, debo proponer traducciones de sus preguntas y encontrar un criterio de éxito para esas traducciones.

Quiero ahora ofrecer un tal criterio de la traducibilidad de unas preguntas a otras preguntas. Su base intuitiva es nuestra sospecha de que, si tenemos dos conjuntos de preguntas, que en ambos casos se contestan atendiendo al mismo tipo de datos, y si el cambio en la respuesta a una pregunta del primer conjunto se correlaciona siempre con un cambio en la respuesta a una del segundo, entonces puede que uno de los conjuntos nos sobre. La mejor ilustración de esta sospecha quizá sea la que nos provocan las explicaciones ad hoc. Considérense las dos preguntas: «¿el opio hace dormir a la gente?» y «¿tiene el opio una virtud dormitiva?» Nuestras respuestas a ambas las determinarán los mismos datos, y si cambiamos nuestra respuesta a una, cambiamos simultáneamente la respuesta a la otra. Al ver que esto ocurre, empezamos a sospechar que una no pregunta más que lo que pregunta la otra y que una de ellas es superflua.

Una vez esbozada esta base intuitiva, voy a proponer el siguiente criterio para lo que llamaré la *traducibilidad heurística* de preguntas. Diré que una pregunta Q es heurísticamente traducible a otra pregunta Q' para un usuario dado si, y solo si:

Para cualquier enunciado S que el usuario acepta como una respuesta a Q existe

(1) un enunciado S_e (donde «e» se refiere a «evidencia») que él aceptaría como evidencia adecuada de S, y
(2) un enunciado S' para el que él también aceptaría S_e como evidencia adecuada y el cual aceptaría como una respuesta a Q'.

Este concepto de «traducibilidad heurística de Q a Q' para un usuario dado» puede generalizarse en un concepto general de traducibilidad heurística del mismo modo en que se generalizó el concepto de traducibilidad evidencial de S a S' para un usuario dado. Podemos decir que una pregunta Q es heurísticamente traducible si, y solo si, para todo usuario del lenguaje existe alguna pregunta Q' (no necesariamente la misma para cada usuario) a la que Q es heurísticamente traducible para ese usuario. Si, para todas las Q's

relevantes, ninguna de las diversas preguntas Q' a las que Q es heurísticamente traducible para los diversos usuarios contiene términos que sean miembros del vocabulario V, entonces podemos decir que el vocabulario V, en su uso como vocabulario temático, es eliminable. Si su uso como vocabulario temático era el primario, entonces podemos afirmar que hemos realizado una reducción radical de ese vocabulario y que hemos demostrado que no hay K's, donde K es el tipo de cosa sobre la que el vocabulario pretendía formular preguntas.

Cuando tal afirmación se interpreta en la forma sugerida por esta definición de la traducibilidad heurística, se puede entender como la afirmación de que ya no necesitamos tener una teoría de K, pues todos los elementos de evidencia que podrían usarse en apoyo de los enunciados que conforman una teoría de K ya han cumplido con su función, por así decir, contribuyendo a responder a las diversas preguntas Q' a las que Q es heurísticamente traducible. Por consiguiente, cuando decimos que, si tenemos la física y la semántica, ya no necesitamos la ontología tomista, o que, si tenemos la psicología y la semántica, no necesitamos lo que Heidegger llama «ontología fundamental», la mejor forma de interpretar lo que afirmamos es que podríamos suministrar traducciones heurísticas de todas las preguntas que pretenden responder la ontología tomista o la ontología fundamental heideggeriana. Esta afirmación, quizá junto con la de que todos los enunciados que hacen las teorías ontológicas de Tomás de Aquino y de Heidegger son evidencialmente traducibles, es lo que se necesita para respaldar nuestra afirmación de que esas preguntas son verbales y, por consiguiente, podemos dejar de hacerlas sin peligro.

Pienso que es ese análisis reductivo de un vocabulario filosófico técnico, que se basa en el intento de suministrar traducciones evidenciales y heurísticas más que equivalencias, lo que de hecho les estamos presentando a nuestros colegas filósofos cuando discutimos con ellos quién creemos que está haciendo las preguntas equivocadas. Al proponer estas definiciones, solo he intentado describir los verdaderos procedimientos que empleamos en tales discusiones. No pretendo haber descubierto ningún argumento

nuevo e imbatible que utilizar en semejantes situaciones. Pero confío en que lo que he dicho baste para justificar mi afirmación, en la sección II, de que el análisis reductivo de vocabularios temáticos, a pesar de su dificultad, es un método viable para discutir qué preguntas necesitan responderse y cuáles no.

---

[1] J. O. Urmson, *Philosophical Analysis: Its Developments between the Two World Wars.* Oxford, Clarendon, 1956, p. 149 [hay trad. esp.: *El análisis filosófico*, Barcelona, Ed. Ariel, 1979].

[2] Ernest Nagel, «The Meaning of Reduction in Natural Sciences», en *Philosophy of Science*, ed. por Arthur C. Danto y Sidney Morgenbesser. Nueva York, Meridian, 1960, p. 308.

[3] Urmson, *Philosophical Analysis*, pp. 183 y 161.

[4] Paul Ziff, *Semantic Analysis*. Ithaca, NY, Cornell University Press, 1960, p. 197.

[5] Urmson, *Philosophical Analysis*, p. 183.

[6] *Ibid*.

[7] No hemos sido capaces de localizar esta cita en *Philosophical Analysis* ni en ningún otro trabajo de Urmson. —Eds.

[8] *Ibid*. [el añadido entre corchetes es de Rorty. —Eds.]

[9] Si mantenemos las palabras-número corrientes, podemos tratarlos como grupos de unidades invisibles y, por consiguiente, discutir sobre el modo de subsistir de los números, su inmaterialidad, su conexión con la intuición pura de la temporalidad, etcétera. En cambio, si adoptamos los sustitutos russellianos de las palabras-número, la discusión de esas cuestiones deja de tener sentido.

[10] Entiendo por una pregunta meramente verbal aquella que nos sentimos dispensados de contestar sobre la base de una simple decisión sobre cómo

usar los términos en el futuro. Sentimos tal dispensa solo cuando nos hemos asegurado de que el propósito primario al que sirve el vocabulario en cuestión no se verá afectado por cualquier decisión que pudiéramos tomar en ese sentido.

[11] Urmson, *Philosophical Analysis*, p. 183.

[12] *Ibid*.

CAPÍTULO 6

FENOMENOLOGÍA, ANÁLISIS LINGÜÍSTICO Y CARTESIANISMO: COMENTARIOS SOBRE RICOEUR

El enfrentamiento entre la fenomenología y el análisis lingüístico es una pugna entre dos grupos de adversarios frente a un único enemigo. El enemigo es el cartesianismo, y cada contrincante intenta sobrepujar al otro cavando cada vez más hondo en busca de supuestos cartesianos ocultos. Ambos se acusan entre sí de ser todavía esclavos de esos supuestos. A los analistas lingüísticos, las descripciones filosóficas del *Lebenswelt* les parecen versiones disfrazadas del tipo de «metafísica revisionista» que practicaban, por ejemplo, Bradley y Whitehead. Esas descripciones, dicen los analistas, solo podrían tomarlas en serio personas que aún no han hecho la terapia necesaria para darse cuenta de que los problemas cartesianos son irreales. A los fenomenólogos, los analistas lingüísticos se les antojan vástagos de aquellos «naturalistas» a los que atacaba el Husserl temprano; parecen los representantes, puestos a la moda de hoy, del «objetivismo» que Merleau-Ponty quería que superáramos. El naturalismo y el objetivismo, dicen los fenomenólogos, solo los pueden seguir tomando en serio quienes estén tan profundamente imbuidos de dualismo cartesiano que no comprendan la necesidad de desarrollar categorías filosóficas nuevas y no-cartesianas.

En lo que sigue, quiero presentar una parte del argumento de los analistas lingüísticos para afirmar que son ellos, y no los fenomenólogos, quienes componen la vanguardia de la revolución anti-cartesiana. Yo mismo no tengo seguro hasta dónde es bueno el argumento, pero parece útil desarrollarlo para ver qué cuestiones hay que solventar antes de poder ponernos de acuerdo sobre qué hace falta que haga la filosofía del lenguaje, si es que tiene que hacer algo. Empezaré presentando tres tesis, como tablas de una

posible plataforma metafilosófica para el analista lingüístico. Después, discutiré en qué medida esas tesis implican o vienen implicadas por la adopción de perspectivas wittgensteinianas sobre el lenguaje. Por último, examinaré ciertas afirmaciones de Paul Ricoeur y las criticaré sobre la base de esa plataforma.[1]

La primera tesis es que la única justificación de que un filósofo redescriba un fenómeno —por ejemplo, pensar, percibir o usar el lenguaje— es que la redescripción haga imposible formular algunos problemas filosóficos tradicionales que el fenómeno había suscitado. Con redescribir me refiero a usar una jerga (que puede suponer acuñar términos técnicos o asignar significados técnicos a los ya familiares) con la que el no filósofo no está familiarizado. Se observará que esta tabla de la plataforma deja totalmente abierto el debate entre los constructivistas como Carnap y los analistas del lenguaje ordinario que se suman a Wittgenstein en pensar que la filosofía «lo deja todo como está». No significa que los filósofos no tengan que usar jerga si es preciso, sino solo que la única excusa que cabe para ella es su utilidad para disolver problemas. Llamaré a esto la tesis Pragmatista.

La segunda tesis es que el único interés que el lenguaje tiene para el filósofo es que descubrir o inventar usos de expresiones lingüísticas le permite disolver problemas tradicionales. Parte del importe de esta tesis es que sobre el lenguaje no hay preguntas trascendentales que hacer, o, por decirlo de otra manera, que todas las preguntas acerca de cómo es posible el lenguaje como tal deben derivarse a disciplinas empíricas (por ejemplo, la psicología, la sociología, la fisiología). A esto lo llamaré la tesis Naturalista.

La tercera tesis es que, si el que una proposición dada exprese una proposición necesaria no se decide apelando al uso lingüístico, nada lo decidirá. Esta tesis es una variante sofisticada de la conocida afirmación positivista de que las verdades necesarias son verdaderas por convención. Nótese que la tesis deja abierto el debate entre los amigos y los enemigos de la distinción analítico-sintético. Dice meramente que, si podemos darle un sentido a la noción de verdad necesaria, ese sentido tendrá que venir explicado únicamente por

referencia a conceptos que aparecen en la metateoría de la lingüística empírica, es decir, conceptos sintácticos y semánticos. Llamaré a esto la tesis Convencionalista.

Wittgenstein adopta de manera bastante explícita en las *Investigaciones* las tesis Naturalista y Convencionalista. No adopta la Pragmatista, sino otra más fuerte, la de que no hay *ninguna* justificación para que el filósofo redescriba fenómenos. Su argumento para ello, hasta donde tiene uno, es que todos los problemas filosóficos tradicionales se pueden disolver simplemente «reconduciendo las palabras de su empleo metafísico a su empleo cotidiano».[2] No quiero defender ni examinar esta última tesis, ya que no creo que esas palabras describan apropiadamente los argumentos anti-cartesianos de Wittgenstein mismo. Más en general, pienso que las observaciones metafilosóficas de Wittgenstein en las *Investigaciones* no concuerdan con sus propias prácticas filosóficas, y que pueden dejarse a un lado sin peligro. Parece que viene más al caso centrar la atención en las conexiones entre su modo anti-cartesiano de ver las relaciones entre lenguaje, pensamiento y realidad, por un lado, y las tesis Naturalista y Convencionalista, por otro.

Wittgenstein arguye desde la premisa (aceptada también por Husserl) de que no hay pensamiento que no esté en palabras, a la conclusión de que no hay forma de contestar a preguntas sobre si nuestro lenguaje es una representación adecuada de la realidad y de qué manera la representa. El argumento, muy a grandes rasgos, es que siempre que queremos comparar el lenguaje con la realidad, nos vemos comparando dos representaciones lingüísticas de la realidad. Uno podría pensar que al menos cabe comparar nuestro lenguaje con la experiencia, pero aquí Wittgenstein insiste en que lo más que podemos hacer es comparar otros trozos de lenguaje con aquellos que son informes de experiencia. Considera la noción de «experiencia no mediada por el lenguaje» o de «experiencia prelingüística» (cuando se interpreta que esta expresión se refiere a experiencias inadecuadamente descritas por el lenguaje cotidiano) como una variante más de la cosa-en-sí kantiana, una entidad

mitológica generada por el uso de un modelo epistemológico erróneo.

No obstante, podría parecer que cabe conceder todo esto y, aun así, insistir en que hay estructuras intencionales que examinar; por ejemplo, las estructuras que subyacen a nuestro uso de signos. Admitiendo que no hay tal cosa como lo «Dado» prelingüístico, podría uno decir, parecería que queda el problema de cómo nos las arreglamos para salvar el salto entre los símbolos lingüísticos y la realidad no lingüística simbolizada. Así, vemos decir a Ricoeur que «estamos separados por siempre de la vida por la función misma del signo».[3] Desde el punto de vista de Wittgenstein, sin embargo, la idea de que existe tal salto o separación es una reliquia más del dualismo cartesiano. Además, la idea de que hay una disciplina cuyo cometido es descubrir las estructuras intencionales que salvan el salto, o bien (a) es solo un modo de describir en jerga la práctica del análisis lingüístico, a saber, detectar entre usos lingüísticos relaciones parecidas a reglas, o bien (b) es una esperanza vana de descubrir rasgos de una mítica experiencia prelingüística. Para hacer valer este dilema, un wittgensteiniano recurrirá a la tesis Convencionalista y argüirá que las únicas verdades necesarias que podemos esperar descubrir son las que están incrustadas en nuestras prácticas lingüísticas. La tesis Convencionalista viene así en auxilio de la Naturalista ayudando al wittgensteiniano a desacreditar pretendidos enfoques trascendentales de los problemas filosóficos y, en particular, de los problemas de filosofía del lenguaje.

Pero en este punto el fenomenólogo podría argüir que el razonamiento wittgensteiniano es circular. Nadie que no haya aceptado ya la tesis Naturalista, cabría decir, aceptaría la tesis Convencionalista. Porque, si la tesis Naturalista es falsa —si hay relaciones necesarias de presuposición entre nuestras prácticas lingüísticas y algo más (llamémoslo la estructura del *Lebenswelt*)—, entonces la Convencionalista es falsa también. Creo que esta acusación de circularidad está bien fundada, pero el cuento no acaba aquí. Porque ahora el wittgensteiniano puede apoyarse en la tesis Pragmatista.

Lo puede hacer del modo siguiente: concedamos, diría él, que hay problemas filosóficos sobre cómo conseguimos usar signos para representar. Lo que queremos es encontrar una forma de describir los signos y sus usos que evite esos problemas, una forma de describir los fenómenos del uso lingüístico en la que no aparezcan «saltos» o «separaciones». Si uno decide llamar a esa descripción una descripción del *Lebenswelt*, bien está. Lo único que importa es que los problemas se eviten. Si, por ejemplo, decimos con Merleau-Ponty que el lenguaje no «expresa pensamientos», sino que «es una toma de posición del sujeto en el mundo de sus significados»[4], eso nos puede ayudar a disolver los pseudoproblemas tradicionales sobre la relación entre un pensamiento y su ropaje lingüístico. Podría ser de mucha más ayuda que, digamos, el ataque cuasi-conductista de Wittgenstein a la idea de los pensamientos como episodios internos, intuibles. Ya he admitido que el propio trabajo de Wittgenstein es más una redescripción de los fenómenos que una simple inspección de los juegos de lenguaje que realmente desplegamos, por ejemplo con palabras como significado y «signo». Gustosamente admito que además puede encontrarse una redescripción mejor.

Pero esta concesión sigue dejando al wittgensteiniano y al fenomenólogo a mucha distancia entre sí. Para el primero, el objetivo de tales redescripciones no es describir con exactitud algo no familiar (la «experiencia vivida», las «estructuras intencionales»), sino lisa y llanamente disolver problemas. Entre decir «ya no tengo los problemas tradicionales porque sé que ahora por fin he descrito con exactitud el *Lebenswelt*» y decir «sé que he descrito con exactitud el *Lebenswelt* porque ya no tengo los problemas tradicionales» media todo un mundo. El wittgensteiniano, presintiendo que el fenomenólogo no dispone de más criterio para una descripción fenomenológica exitosa que el éxito en disolver problemas, se inclina a ver esas referencias al *Lebenswelt* como meros cumplidos que el fenomenólogo se dedica a sí mismo y no hace caso de ellas.

Hasta aquí lo que tiene que ver con la confrontación general entre aproximaciones wittgensteinianas y fenomenológicas a los problemas de filosofía del lenguaje. Quiero pasar ahora a dos cosas que dice Ricoeur y mostrar qué aspecto tienen cuando se miran desde la perspectiva que he estado desarrollando.

Empiezo por lo que dice sobre el preguntar regresivo de Husserl. A Ricoeur le hallamos en guardia contra la acusación de que Husserl recae en nociones cartesianas del «estar dado». Nos previene contra «cualquier recurso a algo parecido a una "impresión" en sentido humeano».[5] Inmediatamente después de esta advertencia, prosigue diciendo que es «desde dentro del mundo de los signos y sobre la base de la *doxa* en el sentido del *Teeteto* —es decir, sobre la base de juicios de percepción— como "indagamos" regresivamente hacia una experiencia vivida primordial; pero esta así llamada experiencia "vivida", para hombres nacidos como nosotros entre palabras no será nunca la presencia desnuda de un absoluto, sino que rehará aquello hacia lo que apunta ese preguntar regresivo».[6] No puedo detenerme aquí a combatir la dudosa insinuación de que, para otros seres que no fueran «hombres nacidos como nosotros entre palabras», podría haber algo como «la presencia desnuda de un absoluto». En vez de eso, quiero preguntar: si de verdad estamos rehaciendo nuestra experiencia vivida a medida que practicamos nuestro preguntar regresivo (o, más en general, como ha dicho Merleau-Ponty, si «el mundo fenomenológico no es el traer a expresión explícita un ser-preexistente, sino el establecer el ser»[7]), ¿en qué sentido puede concebirse la indagación regresiva como un descubrimiento y no como una invención? ¿Cómo sabemos si una experiencia vivida rehecha fue en absoluto vivida alguna vez? Esta noción de una experiencia vivida prelingüística, ¿es algo más que el último disfraz de la cosa-en-sí, algo de lo que nada se puede decir pero que se invoca como recurso heurístico útil para ayudar a explicar qué hacemos cuando hacemos fenomenología? Para decidir si es más que eso, hay que preguntar si tenemos criterios para elegir un conjunto de respuestas a una tras otra de sucesivas *Rückfrage*. No conozco la respuesta a esta pregunta.

En segundo lugar, paso a la insinuación de Ricoeur de que la concepción de Wittgenstein del lenguaje es unidimensional. La insinuación es bastante cierta si significa que Wittgenstein piensa que el lenguaje-como-*lekton*, un *lekton* en el que «el significado es un término dentro de un sistema de dependencias internas»[8], puede reducirse al lenguaje como acto-de-habla. El wittgenstenianismo en filosofía sí conduce a la idea de que la investigación empírica sobre los actos-de-habla de las personas es una base suficiente, y la única evidencia posible, para dar cuenta de esas relaciones sintácticas que componen el «sistema de dependencias internas». (Esto es verdad ya tenga uno un enfoque chomskiano o skinneriano de la lingüística.) Semejante reducción puede fracasar, y entonces peor para Wittgenstein. Pero, entre tanto, no es una crítica del programa reduccionista propuesto el decir que esa concepción del lenguaje es «adialéctica». Estaba pensada para ser eso precisamente. Ni tampoco hace falta ver el lenguaje como «una mediación [...] entre *Logos* y *Bios*» para justificar la afirmación de que «una crítica del lenguaje ordinario es ella misma posible».[9] Podemos adoptar la perspectiva de Wittgenstein de que, para la experiencia humana hasta donde la indagación filosófica la puede alcanzar y describir, no hay salto entre *Logos* y *Bios*, y aun así esperar, *pace* Wittgenstein, transformar el *Bios* mediante la transformación del *Logos*. No hay ninguna conexión necesaria entre la idea de que los signos salvan misteriosos saltos cartesianos y la de que la filosofía puede convertirse en una crítica de la multitud de interpretaciones que depositamos sobre nuestra vida. Si queremos desarrollar una hermenéutica como la que Ricoeur concibe en su *De l'interprétation*[10], podemos hacerlo igual de bien desde un punto de partida wittgensteiniano que desde uno husserliano.

---

[1] Este artículo es una respuesta a «Husserl and Wittgenstein on Language», de Ricoeur, que fue publicado más tarde en *Phenomenology and Existencialism*, ed. por Edward N. Lee y Maurice Mandelbaum. Baltimore,

Maryland: Johns Hopkins University Press, 1967; y en *Analytic Philosophy and Phenomenology*, ed. por Harold A. Durfee. La Haya: Nijhoff, 1976. —Eds.

[2] Ludwig Wittgenstein, *Investigaciones filosóficas*, § 116. [N. del T.: seguimos la traducción al castellano de A. García Suárez y U. Moulines, Barcelona, UNAM / Editorial Crítica, 1988.]

[3] Paul Ricoeur, «Husserl and Wittgenstein on Language», en *Analytic Philosophy and Phenomenology*, ed. por Harold A. Durfee. La Haya: Nijhoff, 1976, p. 95. —Eds.

[4] Maurice Merleau-Ponty, *Phenomenology of Perception*, trad. de Donald A. Landes. Nueva York: Routledge, 2005, p. 225. —Eds.

[5] Ricoeur, «Husserl and Wittgenstein on Language», p. 92.

[6] *Ibid*.

[7] Merleau-Ponty, *Phenomenology of Perception*, p. xxii.

[8] Ricoeur, «Husserl and Wittgenstein on Language», p. 94.

[9] *Ibid*., p. 95.

[10] Paul Ricoeur, *De l'interprétation*, colección «L'ordre philosophique». París: Éditions du Seuil, 1965.

CAPÍTULO 7

LA INCOMUNICABILIDAD DE LAS «CUALIDADES SENTIDAS»

Buena parte de la discusión en torno a la existencia o la naturaleza de los episodios mentales privados generada por las *Investigaciones filosóficas*[1] de Wittgenstein ha girado sobre la tesis de que

(1) no podemos comunicar a otros ciertas cualidades (las «cualidades especiales sentidas») de nuestras sensaciones.

Wittgenstein, o por lo menos algunos wittgensteinianos, la han utilizado como premisa clave en argumentos con conclusiones como que «"dolor de muelas" no es el nombre de una sensación»; «no tiene sentido decir que sé que tengo dolor, porque lo siento»; y «si se construye la gramática de la expresión de la sensación según el modelo de "objeto y designación", entonces el objeto cae fuera de consideración por irrelevante».[2] En este trabajo no quiero rastrear tales argumentos, sino concentrarme en la citada tesis, a la que llamaré la Tesis de la Incomunicabilidad. Quiero aducir que, en un sentido, es verdadera, pero filosóficamente inocua, mientras que en otro es falsa. Esta desambiguación se basará en diferenciar dos sentidos de «conocimiento», y terminaré el artículo arguyendo que una de las raíces del escepticismo epistemológico es no tener en cuenta esa diferenciación.

¿Qué razón hay para pensar que

> no podemos comunicar a otros ciertas cualidades (las «cualidades especiales sentidas») de nuestros dolores?

Obviamente, ningún argumento para esta tesis —la Tesis de la Incomunicabilidad— puede consistir en hacer un listado de tales cualidades, pues, *ex hypothesi*, no disponemos de ningún lenguaje

con el que formular esa lista. ¿La idea de que dichas cualidades existen tiene apoyo en el sentido común? A mi juicio, solo el de que a veces decimos cosas como:

(A) Tú no puedes saber cómo es esa experiencia si no la has tenido nunca.
(B) Tú (que has nacido sin dientes) no puedes saber cómo es tener dolor de muelas (o: …no puedes saber cómo se siente un dolor de muelas).
(C) En aquel momento sentí algo que no encuentro palabras para expresar, pero si se repitiera lo volvería a saber al momento.

Quiero sugerir que «saber» significa en estas tres oraciones «imaginar», «imaginar» en el sentido de «ser capaz de formar una imagen de, o de tener un recuerdo de». Interpretadas así, todos podemos pensar en situaciones donde esas oraciones serían verdaderas. También podemos ver, creo, cómo esas situaciones dan lugar a la idea de «cualidades especiales sentidas». Obviamente, hay algunos rasgos de algunas de mis experiencias (por ejemplo, que ocurrió a mediodía de un miércoles, que tenía algo que ver con mis muelas) que *puedo* comunicar, en el sentido de que tú ciertamente puedes imaginar que tuvieras una experiencia de la cual dirías que tenía los rasgos que yo he mencionado. Pero si, como aseveran (A), (B) y (C), hay un residuo incomunicable (en el caso de algunas experiencias, con respecto a algunos interlocutores), presumiblemente se trata de rasgos de la experiencia que yo *sentí*. Pues ¿cómo sabría yo que hay un residuo incomunicable si *en algún sentido* no supiera qué era ese residuo? ¿Y cómo podría saberlo sino por haberlo sentido? (Después de todo, nadie podía haberme hablado de él.) ¿Y qué podría haber sentido sino algunas cualidades de la experiencia?

Creo que la respuesta correcta a «¿cómo sabría yo que había un residuo incomunicable?» es sencillamente: porque has encontrado que tu discurso es incapaz de infundir la capacidad de *imaginar* algo, donde (repito) imaginar significa «tener imágenes, o recuerdos, o respuestas apropiadas a las apariciones de la cosa en cuestión». Yo

solo sé que no he acertado a comunicar cómo es una experiencia si mi auditorio no responde del modo deseado. Las experiencias no vienen con la etiqueta «incomunicable», ni tampoco ninguno de sus rasgos. La incomunicabilidad es algo que les atribuimos en virtud de no haber acertado a comunicar. El no acertar a comunicar lo comprobamos con preguntas como «¿puedes visualizarlo?» y «¿has tenido alguna vez el tipo de experiencia que haría que hicieras esto o lo otro (olvidarte de lo que te rodea, tener la tentación de postrarte ante la Naturaleza Universal)?» Sabemos que hemos tenido una experiencia única, o por lo menos que nuestro auditorio no nos «sigue», por las respuestas que obtenemos a tales preguntas. No sabemos en ningún sentido qué era ese residuo, sabemos simplemente que había un residuo. Por consiguiente, las «cualidades especiales sentidas» no son objetos de conocimiento. Son modos de resumir incapacidades de comunicar.

Pero podría objetarse: lo que no estoy acertando a comunicar es cierta porción de conocimiento que tengo. Por tanto, debo saber *algo*. Para responder a esta objeción, podemos empezar considerando la distinción siguiente entre dos sentidos de «saber qué es X». En el primer sentido,

> S sabe$_1$ qué es X sii S puede tener a voluntad una imagen o un recuerdo de X y puede reconocer X's a simple vista (o «inmediatamente» por vía de algún otro sentido).

En el segundo caso,

> S sabe$_2$ qué es X sii S puede usar la palabra «X» en la mayor parte de los usos normales (aunque no necesariamente en informes de experiencia) y puede aducir bases adecuadas para creer un gran número de proposiciones verdaderas e interesantes que él expresa mediante oraciones que contienen «X».

Lo relevante de esta discrepancia de significado para el asunto que nos ocupa es que, si «saber» tiene el primer sentido, entonces es ciertamente conocimiento lo que no soy capaz de comunicar en situaciones que piden enunciados como (A), (B) y (C). Porque si

«puede imaginar cómo es X» implica «es capaz de tener a voluntad una imagen o un recuerdo de X», por supuesto que mi superior capacidad de imaginar y reconocer sirve como prueba de que yo sé$_1$ algo que tú no. Pero si «saber» tiene el segundo sentido, entonces eso que no soy capaz de comunicar no es conocimiento, pues yo puedo saber$_2$ muchísimo sobre cosas que ni puedo reconocer a simple vista (o mediante el uso «no mediato» de algún otro sentido), ni de las que puedo formar una imagen o tener recuerdos. Así que ahora nuestra pregunta se convierte en: ¿qué cuenta como saber, por ejemplo, qué significa «rojez» o «dolor»: saber$_1$ lo que es la rojez o el dolor o saber$_2$ lo que es la rojez o el dolor? Naturalmente, eso se reduce a preguntas del tipo: «¿el ciego de nacimiento, por ejemplo, quiere decir con «rojo» lo que quiere decir el vidente?», ¿el hombre que nunca ha sentido dolor, por ejemplo, sabe lo que significa «dolor»?

Si la contestación a estas últimas preguntas es «sí», entonces estamos en el buen camino para bloquear cualquier consecuencia paradójica de combinarlo con la Tesis de la Incomunicabilidad. Porque ahora (A), (B) y (C) solo son verdaderas en el primer sentido de «saber». La Tesis en ese caso no va más allá del hecho de que no puedo dotarte de ciertas capacidades apropiadas de imaginar y reconocer usando el lenguaje. Pero ahora las «cualidades especiales sentidas» de la Tesis dejan de parecer siquiera «objetos de conocimiento». Porque «no puedo comunicarte las cualidades especiales sentidas de mis sensaciones» resulta no significar otra cosa que «no puedo, usando el lenguaje, infundirte la capacidad de tener el tipo de imágenes que yo tengo o de alcanzar las cotas para reconocer cosas que yo alcanzo». Además, *los criterios de tu incapacidad para tener esas imágenes o de alcanzar esas cotas se pueden expresar ellos mismos en nuestro lenguaje compartido y público*. Así, si le digo a un ciego de nacimiento que no puede reconocer objetos rojos a simple vista, él sabe exactamente lo que quiero decir, como también cuando le digo que no puede formar una imagen visual de cosas rojas. Él sabe exactamente qué es lo que no puede hacer, y puede admitir encantado que no puede captar, o

percatarse de, o conocer las cualidades especiales sentidas de la experiencia de ver cosas rojas.

Pero, al admitir esto último, no estará admitiendo que no sabe$_2$ qué es el rojo. Si se le dice que no hay seguridad de que con «rojo» quiera decir lo que nosotros, lo que él debe cuestionar no es (1) —la Tesis de la Incomunicabilidad—, sino

(2) yo sé lo que es un determinado tipo de sensación solo porque sé de ciertas "cualidades especiales sentidas" incomunicables que son características de ciertas sensaciones que yo mismo tengo,

y

(3) yo no sé si sería apropiado aplicar el nombre de un determinado tipo de sensación —por ejemplo, «dolor»— a menos que sepa si estoy hablando de algo que tiene ciertas «cualidades especiales sentidas» incomunicables.

Si aceptamos estas tesis, tenemos que conceder que un hombre que no tiene ciertas capacidades no sabe lo que significa «dolor», y enunciados análogos a (2) y (3) nos forzarían a extraer parecida conclusión negativa respecto de la captación de «rojo» por parte del ciego. Esas tesis expresan la opinión de que, si una clase de cosas tiene cualidades especiales sentidas —esto es, si hay algunas capacidades del tipo relevante (de imaginar, de reconocer) que no puedo infundir en otros usando meramente el lenguaje (sin acompañarlo de ostensión)—, entonces el significado del nombre de esa clase de cosas no se les puede enseñar a aquellos en quienes no pueden infundirse dichas capacidades de esa manera. Pero esta opinión es absurda. Porque cualquier evidencia que se citara en su favor estaría basada en la conducta lingüística de personas que carecerían de esas capacidades, una conducta lingüística que solo valdría como evidencia si tales personas sí conocieran de hecho el significado del sustantivo en cuestión. No podemos saber si un

hombre percibe las superficies coloreadas como coloreadas a menos que sepamos si conoce los predicados-de-color.

El sentido de lo dicho en los últimos párrafos puede resumirse así: referirse a cualidades incomunicables de una experiencia es un modo de referirse a una limitación de capacidades por parte de quienes carecen de la experiencia de hacer ciertas cosas, pero no podemos decir que entre esas cosas esté «entender los términos usados para describir dicha experiencia». Porque si dijéramos eso, no estaríamos en condiciones de saber qué experiencias ha tenido o no ha tenido una persona. Dicho en términos de una crítica a Wittgenstein, no son las sensaciones privadas lo que «cae fuera» o lo que no debe construirse según el modelo de «un objeto que tiene una designación», son simplemente esas cualidades incomunicables. Caen fuera de la cuestión de dominar un lenguaje por irrelevantes. De hecho, decir que es así no es sino decir que lo que cuenta como saber qué es X, donde eso equivale a saber qué significa «X», es saber$_2$ qué es X. Y esto, a su vez, no es más que una forma de decir que el conocer el significado de «X» está garantizado por ser capaz de usar «X» de algunas maneras normales (aunque no necesariamente de todas; por ejemplo, no necesariamente en expresiones de un conocimiento no-inferencial).

Por consiguiente, el lema «el significado es el uso» (rebajado a «si una cantidad razonable de tu uso está bien, conoces el significado») resulta ser el núcleo del argumento contra el escéptico que insinúa que no podemos saber si ambos queremos decir lo mismo con «dolor». Esto cabía esperarlo, y si Wittgenstein se hubiera limitado a invocar ese principio, habría estado sobre terreno firme. Pero, hasta donde yo puedo ver, no lo hizo. En los pasajes sobre las sensaciones privadas de las *Investigaciones*, protestó contra (1) a la vez que daba por sentadas (2) y (3). Así, acabó atacando una serie de proposiciones —por ejemplo, que «dolor» es el nombre de una sensación; que «a veces sé que tengo dolor»— que son perfectamente inofensivas. De ese modo generó un ataque a la entera noción de objetos privados, que ha arrastrado a sus

discípulos a atacar la de acceso privilegiado y otras igualmente respetables e inofensivas, sin que se gane nada con ello.

El meollo de lo que he dicho hasta aquí es que, mientras no tenga uno una idea simplista de cómo adquieren su significado los términos, la noción de unas cualidades incomunicables será filosóficamente inocua. Puede argüirse que el propio Wittgenstein sostenía eso y que han sido solo algunos de sus seguidores quienes han cometido los errores que he sugerido. Pero no deseo entrar aquí en esa cuestión exegética. Una vez que me he escudado en ciertas interpretaciones de Wittgenstein para presentar la distinción entre dos sentidos de «saber qué es X», ahora quiero darle más enjundia introduciendo otras distinciones relacionadas con ella, entre sentidos de otros términos también relacionados, y mostrando cómo se aplican. Considérense las siguientes tesis:

(4) Podemos tener un conocimiento de algo, el cual es diferente a saber que ciertas proposiciones son verdaderas de ello; por tanto, podemos tener un «conocimiento de» que es diferente a un «conocimiento de que».
(5) Podemos tener conocimiento de la verdad de proposiciones que no tenemos medio de expresar en nuestro lenguaje público.

Ha habido un furioso debate entre partidarios y detractores de lo que Sellars llama «nominalismo psicológico» (la idea de que «todo percatarse de algo es un asunto lingüístico»)[3] en relación con la verdad de estas tesis. Pero yo creo que lo único que hay que hacer respecto de este problema es invocar nuestra distinción previa, esta vez en la forma siguiente y más general:

> S sabe$_1$ de algo sii tiene algún conjunto apropiado de disposiciones de conducta no lingüística hacia esa cosa (o representaciones —mentales y físicas— de esa cosa).
>
> S sabe$_2$ de algo sii es capaz de dar evidencia adecuada para sus creencias sobre esa cosa, donde el proceso de dar evidencia implica

> una capacidad lingüística suficiente para hablar inteligentemente sobre esa cosa.

Todo lo que hay que hacer en relación con (5) es invocar la distinción bajo la siguiente forma:

> S sabe$_1$ que *p* sii su conducta no lingüística es la que se esperaría de personas que asentirían a «p», y *p* es verdadero.
>
> S sabe$_2$ que *p* sii *p* es verdadero, él entiende y asiente a «p» y puede proporcionar evidencia adecuada para *p*.

Estas definiciones, con ser toscas, bastan para darnos el sentido en que se dice de animales y niños pequeños que saben de cosas y que saben que algunas proposiciones son verdaderas. La vaguedad de «conjunto apropiado de disposiciones» y «conducta no lingüística que se esperaría» está a la par con la de las bases sobre las que concluimos que los perros saben de los peligros que tienen delante sus amos, que los pájaros saben del norte y el sur y de la dirección de los vientos, que los niños y las amebas conocen la verdad de proposiciones como «a este lado se está más caliente», y que un robot con la dotación apropiada podría saberlo todo sobre todo y conocer la verdad de todas las proposiciones verdaderas. Pero, puesto que son justamente casos así de «conocimiento» los que se citan (y son *todo* lo que se cita) en favor de (4) y (5), no hace falta que nuestro análisis del sentido en que esas tesis son verdaderas sea de grano más fino.

Como esas tesis son trivialmente verdaderas cuando «saber» se interpreta como «saber$_1$», y trivialmente falsas cuando se interpreta como «saber$_2$», ¿por qué tendría que haber un problema con (4) y (5)? ¿Cómo una equivocidad en los sentidos de «saber» ha podido generar perplejidad filosófica? Yo creo que esa perplejidad solo se produce cuando suponemos que el conocimiento de qué son los X's en el segundo sentido debe estar «basado en» el conocimiento de qué son los X's en el primer sentido. Si «basado en» significa «exige como un antecedente causal», la afirmación es normalmente verdadera, pero en ocasiones falsa (como en el caso del

meteorólogo ciego de nacimiento que predice con esmero y exactitud los rasgos visuales de las puestas de sol). No obstante, los filósofos a menudo interpretan «basado en» como «inferido de». Es solo cuando pensamos en una inferencia desde lo que tiene un niño prelingüístico a lo que tiene el meteorólogo, cuando nos topamos con problemas sobre lo privado y lo público. Pero, naturalmente, no puede haber tal inferencia. Las inferencias son de proposición a proposición. La idea de una inferencia a proposiciones desde un determinado patrón de disposiciones a la conducta discriminativa, o desde un determinado patrón de asociaciones o de capacidades de tener ciertos recuerdos o imágenes, no tiene sentido. Ni tampoco puede ser un problema la traducción entre lo que el niño pequeño o el animal saben y lo que sabe el humano adulto. Porque la traducción es de lenguaje a lenguaje, y el lenguaje es justo lo que no tienen el niño y el animal.

El principio que expresa esta noción confusa de inferencia, y que subyace a buena parte de la epistemología tradicional, es el siguiente:

(6) Si «X» es una expresión referencial que también es un término-de-observación, entonces, si no hay seguridad de que en dos personas el patrón de capacidades para reconocer X's mediante observación y para tener imágenes y recuerdos de X's sean similares, no hay seguridad de que el término «X» tenga el mismo significado para esas dos personas.

Este principio da lugar al corolario de que:

(7) Si «X» es una expresión referencial que también es un término-de-observación, entonces nadie que no tenga conocimiento$_1$ de lo que son los X's puede tener conocimiento$_2$ de lo que son los X's.

Es (7) lo que inspira el escepticismo epistemológico sobre el conocimiento del mundo externo. Porque, mientras uno esté convencido de que el flujo de imágenes que atraviesa nuestras

mentes puede ser decisivo para descifrar si tú y yo queremos decir las mismas cosas con lo que decimos (y, a fortiori, conocemos lo mismo), uno se preguntará cómo puede estar seguro de que ese flujo es el mismo. El que ambos lo describamos más o menos en los mismos términos no valdrá para decidir que es el mismo, porque cualquiera que dude de que los dos queremos decir lo mismo cuando aplicamos «rojo» a un granero tendrá las mismas dudas cuando los dos informamos de una imagen mental roja. La hostilidad de muchos filósofos recientes (Gilbert Ryle, por ejemplo) hacia las imágenes mentales —e incluso hacia las «sensaciones en crudo» de Herbert Feigl[4]— se debe en buena medida, creo, al supuesto de que jamás seremos capaces de responder al escéptico epistemológico mientras admitamos que semejantes cosas existen. Pero desestimar algo por irrelevante es tan bueno como negarle existencia. Uno no necesita negar ni la existencia de un sentido perfectamente válido de «saber» en el que puede haber conocimiento (o «percatación», o «consciencia») prelingüístico o no-lingüístico, ni la de particulares mentales no compartibles a los que tenemos acceso privilegiado, para quitarse de preocupaciones sobre la incomunicabilidad de nuestro conocimiento acerca de esos particulares.

---

[1] Ludwig Wittgenstein, *Philosophical Investigations*. Oxford: Basil Blackwell, 1953. [Hay trad. esp.: *Investigaciones filosóficas*. Barcelona, UNAM / Editorial Crítica, 1988.]

[2] *Ibid*., § 293. [N. del T.: seguimos la traducción de A. García Suárez y U. Moulines para UNAM / Editorial Crítica.]

[3] Cf. Wilfrid Sellars, *Science, Perception, and Reality*. Nueva York: Humanities Press, 1963, p. 160. [Hay trad. esp.: *Ciencia, percepción y realidad*. Madrid, Tecnos, 1971.]

[4] Herbert Feigl, *The Mental and the Physical: The Essay and a Postscript*. Minneapolis, University of Minnesota Press, 1967. —Eds.

CAPÍTULO 8

KRIPKE SOBRE LA IDENTIDAD MENTE-CUERPO

Al término de una discusión sobre las ideas de identidad, necesidad y referencia, Saul Kripke traslada algunas de sus conclusiones a la identificación llamada contingente entre sensaciones y estados del cerebro, la cual se ha asociado con algunos materialistas modernos.[1] Él afirma que

(a) podemos distinguir entre términos usados como «designadores rígidos» (aquellos que seleccionarían el mismo objeto en cualquier mundo posible) y términos que no se usan así;
(b) podemos distinguir entre propiedades esenciales y accidentales de los objetos;
(c) si sucede que el objeto seleccionado por dos designadores rígidos «X» e «Y» es el mismo, entonces sucede necesariamente que X = Y, incluso si la identidad no puede conocerse a priori;
(c') los referentes son seleccionados «mediante ciertas propiedades», que pueden ser esenciales o accidentales;
(d) pero en caso de que dos designadores rígidos —«X» e «Y»— seleccionen su referente mediante una diferente propiedad esencial suya, si podemos imaginar algo que tenga una de las propiedades pero no la otra, entonces no puede ocurrir que X = Y, pues eso violaría (c); y
(e) «Estado cerebral» selecciona los estados del cerebro por una propiedad esencial de los estados del cerebro, y «Dolor» selecciona el dolor por una propiedad esencial de los dolores, y ambos términos son designadores rígidos, así que (en virtud de [d]) las dos cosas no pueden ser idénticas.

Aquí el argumento pivota sobre la tesis de que las palabras para «tipos naturales» como «oro», «dolor», «hombre», «estado cerebral», etc. son designadores rígidos[2] que seleccionarían el mismo conjunto de objetos en cualquier mundo posible en el que se refirieran a algo en absoluto.[3] El paradigma de Kripke de designador rígido —el caso con cuya ayuda introduce el concepto— es un nombre propio. Sus ejemplos paradigmáticos de designadores que no son rígidos son términos que significan propiedades accidentales de individuos; así, «el inventor de las bifocales» selecciona (en este mundo) a Franklin en virtud de una de sus propiedades accidentales. Cuando se trata de tipos más que de individuos, la distinción entre designadores rígidos y no rígidos no está tan clara. Se nos dice que «dolor» es un designador rígido y que designa mediante una propiedad esencial de su referente, y que «calor» también lo es, pero que designa mediante una propiedad accidental de su referente, a saber, la de causar en nosotros la sensación S. Se nos dice que «fibra C» puede ser o no un designador rígido (o puede usarse o no como tal).[4]

El criterio para ser un designador rígido es que el término se refiriera al mismo objeto o tipo de objeto en todo mundo posible. No obstante, ese criterio lo deja a uno en la duda sobre la diferencia entre «dolor» o «calor» y «fibra C», o sobre cómo decidir si la expresión «el fenómeno, sea el que fuere, que causa la sensación S» es ella misma rígida lo sea o no «calor». Parece que Kripke quiere una clasificación como mínimo pentapartita: (1) nombres propios que designan rígidamente individuos; (2) términos que designan individuos mediante propiedades accidentales suyas (por ejemplo, «el inventor de las bifocales»); (3) términos que designan rígidamente tipos de cosas mediante propiedades necesarias («dolor»); (4) términos que designan rígidamente tipos de cosas mediante propiedades contingentes («calor»); y (5) términos que designan no rígidamente cosas («fibra C», quizá).

Muy a grandes rasgos, podríamos decir que la tradición a la que Kripke se opone cree que (2)-(5) son más parecidos entre sí que cualquiera de ellos y (1), mientras que él cree que (1) y (3) se

parecen más entre sí que cualquiera de ellos y (2), (5) y quizá (4). (La dificultad de la distinción entre (4) y (5) se discutirá con más detalle más abajo.) Kripke se parece a Aristóteles en pensar que, por ejemplo, «Sócrates» y «hombre» (como términos de lo que Aristóteles llamaba «la categoría de sustancia») tienen en común rasgos importantes que ninguno de ellos comparte con «filósofo» o «maestro de Platón» (términos de las categorías «accidentales»). Por decirlo de otro modo, Kripke seguramente pensaría que tendría algo de raro (¿o de imposible?) usar «este mismísimo inventor de las bifocales» como un designador rígido que selecciona lo mismo en todos los mundos posibles en que las bifocales están inventadas (algo que en un mundo tendría la propiedad de ser un hombre negro; en otro, un hombre blanco llamado Franklin; en otro, un simio; en otro, un autómata, etc.). De manera parecida, Aristóteles creía (por lo menos en la *Metafísica*, si no en las *Categorías*) que una expresión como «esto blanco» (*hod' ho leukos*, donde «*hod' ho*» traduce «este mismísimo») sería de lo más sospechosa, porque solo una sustancia era un «esto» (un *tode ti*).

Para enfocar el conflicto entre esta perspectiva neo-aristotélica y la teoría de la identidad mente-cuerpo, considérese la siguiente línea que podría seguir el teórico de la identidad. Podría empezar concediendo que Kripke tiene razón en distinguir el tipo metafísicamente necesario de identidad que todo tiene consigo mismo de las modalidades definidas en términos de estatuto epistemológico. Pensemos en el tipo de identidad que une a Héspero y Fósforo. Para sorpresa general, ambos resultaron ser el mismo objeto. Así que, con seguridad, Héspero no puede ser otro que Fósforo. La tentación de decir que la identidad de ambos es contingente surge, justo como dice Kripke, de la situación en que imaginamos algo, con algunas de las propiedades accidentales de uno, que carece de algunas de las del otro. Pero el hecho de que las cosas más extrañas puedan resultar ser idénticas entre sí, y de que, cuando eso ocurre, serán vistas como necesariamente idénticas metafísicamente, solo prueba la trivialidad de la noción de identidad metafísicamente necesaria. Una vez que se ha establecido una identidad, es seguro que se verá que es una identidad necesaria.

Pero cuando llamo contingente a la identidad entre dolores y estimulaciones de las fibras C, sigue diciendo el materialista, no estoy negando que este mismísimo dolor es necesariamente idéntico a esta mismísima estimulación. Ciertamente lo es en el sentido trivial en que Héspero es necesariamente idéntico a Fósforo. El argumento de Kripke de que [el dolor y la estimulación] no pueden ser idénticos no pivota sobre la afirmación de que todos los enunciados-de-identidad que vinculan designadores rígidos atribuyen necesidad metafísica, sino sobre la de que la investigación no puede cambiar racionalmente nuestras opiniones sobre qué términos significan propiedades esenciales. Esta última afirmación, si fuera verdadera, significaría que nunca puede haber una reducción filosóficamente sorprendente. En tales reducciones se incluyen los casos donde queremos decir, simplemente sobre la base de la navaja de Ockham, que el pretendido tipo natural A es el mismo que el tipo natural B. Si los A's y los B's están correlacionados constantemente, y si hay conveniencia en que la referencia a B's tome el lugar de la referencia a A's en el patrón global de explicaciones y predicciones, hemos descubierto con ello que cualquier A dado probablemente es necesariamente idéntico metafísicamente a algún B dado. Kripke ve una barrera a esa identidad en el hecho de que podemos imaginar A's que existen sin B's, pero esto demuestra meramente que lo que podemos imaginar es relativo a nuestro estado de conocimiento. Estas cuestiones epistémicas nada tienen que ver con la necesidad metafísica. Lo que Kripke debería decir es que podemos imaginar un mundo donde algunas propiedades de la cosa que es a la vez un A y un B —por ejemplo, la propiedad de A-eidad o la de B-eidad— las tuvieran otras cosas que no son A's o B's. En cambio, su posición da a entender que, en casos de pretendidas identidades de tipos naturales, la única clase de identificación que podría funcionar es lo que a veces se ha llamado la «variante de la desaparición» de la teoría de la identidad, según la cual nunca hubo A's en absoluto, sino solo, por así decir, «A's para los tontos», o sea, «B's».[5] Debido a que Kripke piensa que en algunos casos puede seleccionar mediante experimentos mentales qué propiedades de los A's son necesarias

sin esperar a la ciencia —qué términos están, por así decir, «en la categoría de sustancia», y por ende son esenciales a sus poseedores —, cree que no hay tal cosa como imaginar que las cosas seleccionadas por esas propiedades pudieran resultar no poseerlas después de todo. Sin embargo, si esas cosas de hecho son B's, tiene perfecto sentido que digamos que alguna vez creímos que su esencia era ser A's, pero que la ciencia (y la navaja de Ockham) ha mostrado que estábamos equivocados: su esencia, a lo sumo, es ser B, y ser A es, a lo sumo, solo una propiedad contingente más que resultan tener. Las propiedades esenciales y los tipos naturales discretos se descubren por investigación científica; no son constricciones a ella. Incluso en tiempos de Aristóteles, el esencialismo era una corriente reaccionaria contra los esfuerzos reduccionistas ilustrados y progresistas de platónicos y atomistas. Las intuiciones en que se apoyaba Aristóteles —que algunos términos designan la misma cosa en cualquier mundo posible en que exista, que los enunciados-de-identidad que vinculan a esos designadores, si son verdaderos, lo son necesariamente, y que algunas propiedades son esenciales a las cosas y otras les son accidentales— se pueden preservar, pero no se debe dejar que bloqueen la senda de la investigación. No lo harán si permitimos que los desarrollos científicos ulteriores puedan revelar que términos que alguna vez pensamos que seleccionaban referentes mediante sus propiedades esenciales no lo hacían, y que por tanto dejen de usarse como designadores rígidos. (Aunque, como demostró el abandono gradual de la jerga aristotélica en el siglo XVII, las intuiciones aristotélicas ya no parecen tan importantes una vez se acepta este punto epistemológico.)

Hasta aquí la primera línea de defensa del materialista frente al argumento de Kripke. Sería natural que Kripke replicara señalando que, de hecho, él sí consiente un montón de reducciones «filosóficamente interesantes» —por ejemplo, la del calor a movimiento molecular—, pero que en el caso de las sensaciones hay un rasgo especial que hace que la analogía sea inapropiada. En las reducciones que a los teóricos de la identidad les gusta citar, se usa algún término en lo que Aristóteles llamaría «la categoría de

relación» (por ejemplo, «el fenómeno, sea el que fuere, que es la causa típica de la sensación S») para seleccionar un referente que la investigación finalmente descubre que tiene tal y tal esencia.[6] De este modo, el calor tiene una naturaleza (a saber, el movimiento molecular) que puede descubrirse, pero el dolor no, porque nosotros ya conocemos la esencia del dolor. Así, nos dice Kripke:

> En el caso de la identidad del calor con el movimiento molecular, la consideración importante era que, aunque «calor» es un designador rígido, el referente de ese designador es fijado por una propiedad accidental del referente, a saber, la de producir en nosotros la sensación S. Así, cabe que un fenómeno hubiera sido rígidamente designado de la misma manera como un fenómeno de calor, con su referente seleccionado igualmente mediante la sensación S, sin que ese fenómeno fuera calor y, por consiguiente, sin que fuera movimiento molecular. El dolor, en cambio, es seleccionado por la propiedad misma de ser dolor, por su cualidad fenomenológica inmediata. De modo que el dolor, a diferencia del calor, no solo es rígidamente designado por «dolor», sino que la referencia del designador está determinada por una propiedad esencial del referente.[7]

Kripke explica, presentando el asunto de otra manera, que «la ilusión de contingencia» en la identidad de calor y movimiento molecular la produce la «situación epistémica» creada por el modo en que «se determinan las referencias de los designadores»; así:

> En el caso del movimiento molecular y el calor hay algo, a saber, la sensación de calor, que es un intermediario entre el fenómeno externo y el observador. En el caso de lo mental y lo físico tal intermediario no es posible, pues aquí se supone que el fenómeno físico es idéntico con el fenómeno interno mismo. Alguien puede estar en la misma situación epistémica en la que estaría si hubiera calor incluso si no hay calor, simplemente sintiendo la sensación de calor; e incluso habiendo calor, puede tener la misma evidencia que tendría en ausencia de calor simplemente no teniendo la sensación S. Esa posibilidad no existe en el caso del dolor o el de otros fenómenos mentales. Estar en la misma situación epistémica que se daría si uno tuviera un dolor *es* tener un dolor; estar en la misma situación epistémica que se daría en ausencia de un dolor es no tener un dolor.[8]

Kripke ve tres pretendidas entidades en el caso del calor: la sensación de calor («S»), el calor y el movimiento molecular. La ciencia puede descubrir que las tres son solo dos, ya que «calor = movimiento molecular» es verdadero en virtud del hecho de que «calor», aunque no es sinónimo de «fenómeno que típicamente produce S»[9], usa esta última propiedad «accidental» del calor para seleccionar el calor como su referente. En el caso del dolor ve solo dos entidades, el dolor y los estados cerebrales constantemente correlacionados con él (estimulaciones de fibras C), ambos significados por términos que designan rígidamente y que lo hacen presumiblemente mediante propiedades esenciales de sus referentes. El dolor es seleccionado por «una cualidad fenomenológica inmediata», pero el calor no lo seleccionamos por cómo se siente, sino, por así decir, a distancia, como aquello que nos está haciendo sentir eso. La razón de ser de esta distinción es presumiblemente que puede haber calor incluso si no hay organismos sentientes, pero no dolor[10]; esto es lo que hace «interno» al dolor y «externo» al calor y lo que nos da una «apariencia» de calor (la sensación S), pero ninguna «apariencia» de dolor.

Este último punto sale también a relucir cuando Kripke apela a nuestra intuición de la «esencialidad» para los dolores del *ser un dolor*:

> [El teórico de la identidad] difícilmente puede sortear la dificultad arguyendo que, aunque B no podría existir sin A, *ser un dolor* es meramente una propiedad contingente de A, y que por tanto la presencia de B sin dolor no implica la presencia de B sin A. ¿Puede haber un caso más obvio de esencia que el hecho de que *ser un dolor* es una propiedad necesaria de cada dolor? El teórico de la identidad que quiera adoptar la estrategia en cuestión debe argüir que incluso *ser una sensación* es una propiedad contingente de A, porque *prima facie* parecería lógicamente posible que B pudiera existir sin ninguna sensación con la que plausiblemente pudiera ser identificado. Considera un dolor concreto u otra sensación que una vez tuviste. ¿Te parece en absoluto plausible que *esa mismísima sensación* hubiera podido existir

> sin ser una sensación, a la manera en que un cierto inventor (Franklin) podría haber existido sin ser un inventor?[11]

Convendrá que tomemos primero esta última formulación y desde ahí operemos hacia atrás hasta la idea de que el dolor es seleccionado por «una cualidad fenomenológica inmediata». La comparación con Franklin recalca cómo Kripke está insistiendo en que solo los términos-de-sustancia logran designar rígidamente y solo las sustancias tienen propiedades esenciales. Porque… considérese la siguiente réplica. Concediendo que no podemos imaginar que esta mismísima sensación exista sin ser una sensación, tampoco podemos imaginar que el inventor de las bifocales exista sin ser el inventor de las bifocales. Puedo imaginar que Franklin exista sin ser el Inventor de las Bifocales, pero no que el Inventor de las Bifocales exista en esas condiciones (donde las mayúsculas indican la identidad transmundana de ese personaje, ora encarnado en Franklin, ora en Jefferson, ora en un robot). Está en la esencia de Franklin el ser un hombre, pero no en la esencia del Inventor de las Bifocales. Análogamente, está en la esencia de Dolor (un tipo natural que aparece en cualquier mundo en el que haya fibras C estimuladas) el ser una estimulación de fibras C, pero es contingente que en este mundo, en nuestro tipo de organismo (¿o quizá solo en mí, como afirma el solipsista?), venga junto con una determinada cualidad fenomenológica. Todo depende de en qué pienses como sustrato y en qué como propiedad. Normalmente pensamos en los hombres como sustratos y en sus logros como propiedades, pero podríamos hacerlo al revés. Desde Descartes, las «cualidades fenomenológicas inmediatas» han sido las candidatas primeras a «propiedades esenciales» de ese gran sustrato, la «Mente» (el término que remplazó a «hombre» como paradigma de sustancia a resultas de la revolución científica del siglo XVII). Pero en una cultura materialista, donde por «la mente» se entendiera un resumen de cómo hablaban de sus estados mentales nuestros antepasados, el único sustrato que habría por allí sería el cerebro; para entonces, seguro que la etnología galáctica habría encontrado montones de cualidades fenomenológicas diferentes que aparecen en conjunción

con las estimulaciones de las fibras C. Así que el intento de Kripke de ponernos la piel de gallina diciendo «considera un dolor concreto u otra sensación que una vez tuviste» es una petición de principio: quiere que consideremos como de la esencia la horridez fenomenológica inmediata, pero lo que llega a ser de la esencia no se determina con una vuelta *zu den Sachen*, sino mirando el contexto de discurso sobre la cosa en cuestión. El teórico de la identidad está sugiriendo un contexto nuevo, que él cree que sitúa todo el asunto bajo una nueva luz y tiene diversas ventajas filosóficas. La luz nueva se arroja, como de costumbre, convirtiendo los viejos sustratos en propiedades y las viejas propiedades en sustratos.

La estrategia que, según vengo sugiriendo, tiene a su disposición el teórico de la identidad está expuesta a varias objeciones, siendo la más obvia la de que el sujeto ha sido cambiado de dolor a Dolor, y de Franklin al Inventor de las Bifocales. Pero (como nos recuerda la ambigüedad de *hypokeimenon* entre «sujeto del discurso» y «sujeto de las propiedades») lo que para unos es un sujeto cambiado, para otros es un descubrimiento metafísico. Antes de abordar esas objeciones, permítanme unas pocas observaciones históricas que creo que ayudarán a ver mejor el problema. Por supuesto, el esencialismo de Kripke es un escándalo para los quineanos, para los filósofos del lenguaje que hablan de «conceptos de *cluster*» y para los metafísicos whiteheadianos y bradleyanos. Pero, además, eso que dice de que los dolores se seleccionan por sus cualidades fenomenológicas inmediatas, para los wittgensteinianos es un disparate. Kripke está yendo a contrapelo de prácticamente todas las modas actuales, como debe ser en todo retorno al aristotelismo. Porque gran parte de la filosofía del siglo XX es una reacción contra las ideas aristotélicas de la lógica y del lenguaje[12] y contra los remanentes de aristotelismo en las filosofías cartesianas de la mente.[13] En todos los frentes, el lema común a esa reacción anti-aristotélica es que lo que cuenta como un individuo, o un sustrato, o una esencia, siempre es relativo a una descripción. La elección entre descripciones alternativas, se supone tácitamente, es una cuestión

de lo que más convenga a las necesidades de la ciencia empírica en su marcha hacia adelante, cuando no simplemente de gusto, de interés o de retórica. De este modo, los materialistas se sienten libres de —y, de hecho, obligados a— construir filosofías del lenguaje que harán del mundo filosófico un lugar seguro para el materialismo, fiados de que al hacerlo están depurándonos y depurándose de los últimos vestigios de superstición aristotélica. Las consecuencias contraintuitivas de tales filosofías del lenguaje constituyen el punto de partida de los esfuerzos de Kripke por sistematizar las intuiciones aristotélicas que la filosofía del siglo XVII había intentado reprimir o sublimar.

Antes de reflexionar sobre la vasta pregunta «¿todo esto es relativo sin más a una elección de descripción?» (y la más vasta aún, «¿cómo distinguiría uno si es así?»), consideremos los diversos recursos de que dispone el wittgensteiniano que desconfía en cuanto se mencionan cualidades fenomenológicas inmediatas. Por ejemplo, puede decir que, como no hay una manera clara de asegurar que yo tengo la misma cualidad fenomenológica inmediata que tú cuando ambos usamos «dolor» (o que cuando yo mismo lo usé ayer), no habría que pensar del término «dolor» que tiene un referente y habría que abandonar la entera noción de «cualidad fenomenológica». Este punto de vista de la «sensación-como-escarabajo-en-la-caja», con las sensaciones «quedándose fuera» cuando se reconocen las dificultades epistemológicas, permite al wittgensteiniano unir fuerzas con el quineano, ambos bajo el lema «ninguna entidad sin identidad». Una segunda jugada posible es decir que el término «dolor» sí refiere de hecho, pero que no seleccionamos sus referentes mediante una cualidad fenomenológica inmediatamente dada, sino mediante un patrón de relaciones causales (efectos típicos de los fogones calientes, disposiciones a respingar y gritar, etc. etc.). Esto encajará con una explicación en términos de «rol causal» de los estados mentales en general y abrirá de par en par la puerta a identificaciones teóricas del tipo calor-movimiento molecular. Porque ahora «dolor» seleccionará justamente estimulaciones de fibras C mediante la propiedad accidental de ser un nodo de determinados patrones causales, tal

como sugieren Armstrong y David Lewis. Como tercera jugada posible, puede decir que «dolor» realmente refiere a un individuo seleccionado mediante una cualidad fenomenológica inmediata, pero insistiendo en que la «identificación» de semejante cualidad es solo una cuestión de tener una determinada práctica lingüística, y que no tiene sentido la pregunta «¿a los marcianos, que no han oído hablar de cualidades fenomenológicas inmediatamente dadas y que usan la expresión "fibras C estimuladas" donde nosotros usamos "dolor" (para explicar los espasmos, los gritos, etc.), se les da inmediatamente la misma cualidad fenomenológica?» Desde esta tercera perspectiva, el hecho de que usemos esa cualidad para seleccionar dolores obviamente no es señal de que dicha cualidad esté en la esencia del dolor, pues obviamente no cabe discutir si somos nosotros o los marcianos quienes usan la cualidad «correcta» para seleccionarlo, y así el escéptico sobre las otras mentes no tiene a qué agarrarse. Las tres estrategias presuponen la idea de que es la forma en que hablamos lo que determina la esencia: la primera no nos dejará referirnos a dolores en absoluto, mientras que la segunda y la tercera nos dejarán hacerlo mediante propiedades «accidentales». Solo estas dos últimas congenian con el materialismo (las modalidades de la «traducción» y de la «desaparición» de la teoría de la identidad, respectivamente), pero ambas dan por hecho que podemos etiquetar lo que queramos como propiedad «esencial» o como propiedad «accidental», dependiendo de lo que convenga a nuestros propósitos científicos y filosóficos ulteriores. Las tres vías se precian de deshacerse del Problema Mente-Cuerpo y del Problema de las Otras Mentes, y las tres desarrollan explicaciones de la referencia, el significado, etc., que se ajustan a ese fin terapéutico.[14]

Es de suponer que Kripke se resistiría a todas estas tácticas e insistiría en que «dolor» es el nombre de un tipo natural de particular que lleva su esencia escrita en la frente. Los wittgensteinianos ven en esa insistencia una confusión entre «la lógica de nuestra forma de hablar de objetos físicos» y «la lógica de nuestra forma de hablar de sensaciones».[15] Es difícil saber quién se

confunde aquí, pero una cosa parece clara: todo gira alrededor de la idea de «identidad mediante una propiedad (necesaria) esencial», y no del «principio de la necesidad de las identidades que utilizan designadores rígidos».[16] Este último puede ser aceptado tranquilamente por cualquiera de las tres tácticas wittgensteinianas sugeridas. Pueden convenir encantadas en que la identidad necesaria de este mismísimo dolor con la estimulación de algunas fibras C es tan estrecha como la identidad necesaria de Héspero con Fósforo. Cuando Kripke dice que «el teórico de la identidad no puede admitir la posibilidad [de que el cerebro de Pérez podría haber estado exactamente en ese estado en el momento en cuestión sin que Pérez sintiera dolor alguno] y continuar desde ahí»[17], pueden replicar que no necesitan admitirla, como Kripke no necesita admitir que Héspero podría existir sin Fósforo. Esas tácticas estaban hablando de modalidades epistemológicas; una vez que Kripke cambia a modalidades metafísicas, por supuesto que tiene toda la razón. Lo que tal vez está imaginando Kripke, dicen, es una situación en la que el cerebro de Pérez está exactamente en ese estado, pero Pérez sigue adelante con indiferencia, niega sentir dolor, etc. Habrá docenas de razones, podrían decir, de que Pérez haga eso. Una es que no sabe de sentir dolores: ha sido criado por marcianos y todo lo que sabe es que sus fibras C están siendo estimuladas. (Nótese que Kripke tiende a identificar sentir dolor con estar en una determinada situación epistémica respecto del dolor, insinuación que mezcla de manera dudosa sentir y saber.) Su indiferencia obedece a su masoquismo, a su educación espartana, a que tiene la mente en otras cosas o a lo que tú quieras. O a lo mejor es que Kripke está imaginando a un hombre hecho de tal modo que, cuando su cerebro está exactamente en ese estado, afirma sentir un placer enorme, bailotea, pide que le mantengan en el mismo estado, etc. Aquí todo sigue bien: él continúa sintiendo dolor, solo que es así como él lo siente, a él le da una cualidad fenomenológica inmediata diferente. ¡Pero entonces no es dolor! Desde luego que lo es, exactamente igual que Héspero seguiría siendo Héspero si nunca hubiera aparecido de noche, sino que la gente hubiera creído que lo hacía

por efecto de una combinación de espejos e hipnosis masiva. Si vamos a tratar este mismísimo dolor en la línea de esta mismísima estrella Héspero, entonces no está bien decir que no puede ser dolor si no viene acompañado de cierta conducta, cualidades fenomenológicas inmediatas, etc. (como no lo está decir que no puede ser Héspero si no aparece en el cielo nocturno). ¡Pero el dolor *es* una cualidad fenomenológica inmediata! Bueno, en ese caso es cierto que no es idéntico a la estimulación de fibras C, pero entonces ciertamente no es un tipo natural de particular seleccionado mediante una de sus propiedades esenciales. Las cualidades no tienen propiedades fenomenológicas inmediatas ni propiedades esenciales. ¡Pero esa cualidad fenomenológica terrible es esencial al dolor! Bueno, eso es justamente lo que estamos discutiendo, si lo es o no lo es. Tiene que haber alguna razón independiente de por qué es esencial, aparte de la afirmación de que podemos imaginar que el estado del cerebro de Pérez es exactamente el mismo y Pérez no siente dolor. Porque la afirmación es falsa si esa cualidad fenomenológica es accidental al dolor, igual que es falso que podamos imaginar Héspero sin Fósforo si el aparecer por la noche es accidental a Héspero.

El dilema general al que se enfrenta Kripke cuando se encara con estas estrategias es el siguiente: si los nombres para los tipos naturales son como los nombres propios en que seleccionan perfectamente bien un objeto incluso si no conocemos su naturaleza e incluso si prácticamente todas nuestras creencias sobre él son falsas, entonces nunca va a haber un argumento a priori para decir que dos tipos naturales no pueden resultar ser el mismo tipo natural. Si no son como los nombres propios porque hay ciertos criterios que nos permitirían decir «bueno, no podemos seguir diciendo que la expresión "ese mismísimo K" selecciona esa cosa, ahora que sabemos que nos equivocábamos al pensar que cumplía los criterios de la K-eidad», entonces reaparecen todas las viejas consideraciones epistemológicas sobre conceptos de *cluster* y sobre la relatividad de la esencia al presente estado de la investigación científica. El único modo de pasar entre esos dos cuernos es utilizar la idea de «seleccionar mediante una propiedad esencial» y tener medios a

priori de saber qué propiedades son esenciales. Pero cualquier afirmación a priori de ese tipo va a ser una forma disfrazada de la idea de «criterios para ser un K». Pasar de decir que «dolor» significa «la cosa que tiene esa cualidad fenomenológica inmediata» a decir que seleccionamos el dolor mediante esa cualidad fenomenológica inmediata que le es esencial no cambia nada. En cualquiera de las dos formulaciones, lo que tenemos es una resistencia (¿no poder imaginar?, ¿no querer?, ¿acaso importa?) a usar una determinada palabra si el objeto pretendido no tiene un determinado rasgo. Si hay tal resistencia en el caso del dolor, entonces por eso «ese mismísimo dolor» no es como «Héspero». Porque Héspero podría haber seleccionado cualquier cosa —una estrella, un platillo volante, un dios en un carro de fuego— y no habría sido inimaginable llamarlo «Héspero» una vez hubiéramos descubierto lo que habíamos seleccionado. Al contrario, llamarlo Héspero habría sido tan razonable como siempre fue. Una vez que descubrimos que era una estrella y no un dios, pudimos empezar a decir algo sobre sus propiedades esenciales, pero eso sería porque sabemos algo sobre las propiedades esenciales de las estrellas, no porque sepamos algo sobre qué cosa tiene derecho a ser llamada «Héspero».[18]

Mi conclusión es que amalgamar como propone Kripke términos de sustancias primeras y términos de sustancias segundas, por decirlo así, deja la cuestión sobre la identidad mente-cuerpo como estaba. Las viejas discusiones se traspasarán al nuevo vocabulario, hablando menos de significado y más de referencia, pero sin quebranto dialéctico para ninguna de las partes. El teórico de la identidad tendrá una serie de cosas paradójicas que decir, como siempre, para sostener su posición, y un montón de objeciones a su idea ahora pueden resumirse en la forma «estás confundiendo un accidente de los estados mentales con su esencia» (lo mismo que se dijo del conductismo de Watson hace cincuenta años). Pero el que se haya incurrido en esa confusión no se determinará aludiendo a la necesidad de las identidades que usan designadores rígidos. Ese

principio es lo bastante laxo como para volverlo en favor de los intereses de cualquiera de las partes.

* * *

Ya he argumentado mi tesis principal lo mejor que he podido y quiero terminar con algunas observaciones generales sobre varios tópicos que he tocado, con la esperanza de que rellenen el cuadro que quiero dibujar. Tienen que ver con: (1) la noción de «designador rígido»; (2) la noción de «esencia»; (3) lo que he estado llamando «aristotelismo»; y (4) el materialismo.

Designadores rígidos. Kripke ha definido «designador rígido» como «un término que designa el mismo objeto en todos los mundos posibles». Esta noción es bastante clara para términos que se aplican a individuos o a sus propiedades. Creo que la intuición titubea más cuando pasamos a los tipos naturales. Kripke no argumenta, hasta donde yo veo, que los términos para ellos sean designadores rígidos,[19] y cuesta aclararse sobre lo que viene a decir la afirmación exactamente. Para aclararla necesitaríamos ejemplos de términos que no fueran ni designadores rígidos ni propiedades de individuos. En un lugar afirma que los únicos términos generales que no son designadores rígidos son los que expresan propiedades: por ejemplo, «estúpido», «gordo» y «amarillo».[20] En otro sitio, en cambio, afirma que «caliente», «sonoro» y «rojo» también son designadores rígidos.[21] La idea de seleccionar la misma cosa en cualquier mundo posible (en cualquier situación contrafáctica) no parece darnos nada con lo que avanzar en el caso de «rojo», o «gato», o «dolor», aparte de un conocimiento de las propiedades necesarias de cualquiera de esos tipos de cosa. En el caso de nombres propios, uno puede decir «estoy hablando, da igual cuál sea la situación contrafáctica, de un trigésimo séptimo presidente de EE. UU. reaccionario generalmente llamado "Nixon" y que tiene el aspecto de esta fotografía», o uno puede decir «estoy hablando, da igual cuál sea la situación contrafáctica, de este mismísimo hombre de aquí». Sin duda, Kripke tiene razón al decir que la mayoría de

nosotros usamos normalmente los nombres propios de forma que esta última observación sería apropiada. ¿Pero es posible una elección de intenciones paralela para palabras como «gato», «dolor» y «amarillo»? Podemos conceder que, como dice Kripke, la referencia de esas palabras normalmente queda fijada en términos de la propiedad «ser del tipo ejemplificado por esta muestra»[22], no mediante una especificación de diccionario de condiciones necesarias y suficientes a la manera de Mill.[23] Pero, una vez que la referencia queda así fijada, ¿qué haríamos para distinguir entre seleccionar gatos mediante la idea «cualquier cosa, en cualquier situación contrafáctica, que tenga la siguiente lista de cualidades entre las ejemplificadas por esta muestra» y mediante…? ¿Mediante qué? ¿Mediante si son sencillamente gatos, sea o no sea que se aplique la lista? ¿Pero qué es usar la idea de «gatos, sea o no sea…», salvo que tengamos fabricada alguna otra lista?

Kripke aclara más la noción de «designador rígido» diciendo:

> Cuando digo que una descripción podría haber referido a algo diferente, quiero decir que, en *nuestro* lenguaje tal como *nosotros* lo usamos para describir una situación contrafáctica, podría haber habido un objeto distinto que satisficiera las condiciones descriptivas que *nosotros* damos para la referencia. Así, por ejemplo, usamos la expresión «el inventor de las bifocales», cuando estamos hablando de otro mundo posible o de una situación contrafáctica, para referirnos a quien quiera que, en esa situación contrafáctica, hubiera inventado las bifocales.[24]

Pero esto no ayuda con nuestro problema. Un historiador de la optometría de ideas fijas designa la misma cosa (ora Franklin, ora Jefferson, ora un simio) con «el Inventor de las Bifocales» a medida que inspecciona diversas situaciones contrafácticas. (Si se prefiere un ejemplo más realista, tómese «El Dalai Lama»; para los budistas tibetanos solo ha habido y podría haber uno, da igual la contingencia contrafáctica.) Ya que pensar en Franklin y Jefferson como la misma cosa nos parece contraintuitivo, encontramos raro ese uso del historiador. En el vocabulario de Kripke, eso lo indicamos diciendo que usamos «inventor de las bifocales» no-rígidamente, lo cual es

un modo de expresar nuestra intención de dejar que Franklin y el Jefferson contrafáctico que inventó las bifocales cuenten como cosas diferentes, en vez de como la misma cosa vestida con juegos diferentes de propiedades.

Pero ¿qué se corresponde con este tipo de intención en el caso de los gatos, o los dolores, o la amarillez? En el caso de individuos podemos echar mano prima facie de la distinción sustrato-propiedad (y también, por tanto, de la rigidez de nuestro designador). Pero, en el caso del tipo gato, no parece haber nada de lo que echar mano hasta que uno se haya fabricado la lista de propiedades prominentes (ni, por tanto, manera alguna de comprobar la rigidez). De lo que se puede echar mano, entonces, es de la distinción esencia-accidente. Uno puede discrepar acerca de cómo llamar a la muestra original si todos o la mayoría de sus miembros resultan carecer de lo que uno ha intuido o descubierto ahora que es la esencia del gato. Uno puede llamarlos, como Kripke, «gatos para los tontos»[25] y concluir, dadas algunas otras condiciones, que jamás hubo gatos. O uno puede seguir llamándolos «gatos» y decir que la así llamada intuición o descubrimiento [anterior] de la esencia de ser un gato estaba equivocada del todo. Pero antes de que alguien haya puesto sobre la mesa una teoría de cuál es la esencia, mientras estamos simplemente delante de la muestra original y ni siquiera nos hemos puesto a enumerar cualidades todavía, ¿qué significa decir que, en ese momento, «gato» lo estamos usando «para referirnos a lo que quiera (no necesariamente incluida esta muestra) que, en cualquier situación contrafáctica, sería un gato» (no-rígido), o quizás «para referirnos en cualquier situación contrafáctica a, como mínimo, esas cosas que tenemos delante» (rígido)? El uso por parte de Kripke de la noción de «designadores rígidos» para los tipos naturales parece que quiere abstraerse por completo de la cuestión epistemológica de si hemos captado ya, o creemos haber captado, la esencia del tipo. Pero no veo cómo puede hacerse abstracción de eso.

Conocimiento de propiedades esenciales. La idea de una propiedad de la que una cosa no podría carecer y seguir siendo la misma cosa parece inocua una vez que se la separa de los

problemas epistemológicos. Pero rara vez se la ha separado de ellos, y su impopularidad en la filosofía reciente obedece al hecho de que nadie ha sabido nunca cómo decidir si nuestras corazonadas, o intuiciones, o teorías, sobre qué es esencial a qué son correctas. En consecuencia, la idea de unas propiedades internas ha sucumbido al verificacionismo y ha sufrido la humillación suprema de ser relativizada a «una elección de descripciones» o a «un motivo de investigación». Kripke quiere restablecer su pureza metafísica y, por ende, su respetabilidad. Esto sería bastante inofensivo, solo que él invoca la idea para dirimir disputas metafísicas concretas. En el caso mente-cuerpo, quiere usar la noción epistemológica de «imaginabilidad». Pero con ello se priva a sí mismo de las ventajas de haber hecho de la esencia una noción metafísica, porque ahora se ve arrastrado a peleas entre intuiciones en competencia, entre intuiciones versus necesidades de la ciencia unificada, y todo lo demás. En particular, necesita tener opinión sobre cuándo —si es que alguna vez— puede la investigación ensanchar el radio de la imaginación, cuándo la investigación puede equivocarse sobre la esencia, y cómo lo sabe uno. ¿Podría resultar que el calor no es movimiento molecular? ¿Podría seguir siendo movimiento molecular y resultar, con todo, que no es la causa de la sensación S? Este tipo putnamesco de preguntas con encerrona —las preguntas que, cuando se formulan en términos de significado, dan lugar a la teoría de los «conceptos de *cluster*» que es anatema para Kripke— aparecen también cuando se formulan en términos de «propiedades mediante las que seleccionamos referentes». La opinión quineana de que, cuando las cosas se ponen borrosas, da un poco igual que digamos que estamos hablando de la misma cosa, pero diciendo de ella cosas en su mayoría nuevas, o que ahora estamos hablando de una cosa diferente, conserva toda su fuerza incluso cuando se adopta el «principio de la necesidad de las identidades que usan designadores rígidos». Porque, como he estado arguyendo más arriba, en el caso de las palabras-tipo no se le puede dar un sentido claro a la distinción rígido-versus-no rígido salvo por referencia a la distinción esencia-accidente, y ninguna de las dos puede aplicarse al

problema mente-cuerpo sin verse uno envuelto precisamente en las cuestiones epistemológicas que Kripke quiere evitar.

Aristotelismo. ¿Son entonces «relativas a una descripción» tanto la esencia, lo que cuenta como sustancia y no la cualidad, como la necesidad? ¿Aristóteles se equivocaba sin más, y Quine está en lo cierto con su visión de la necesidad como «solo una gama de grises» y con su desdeñosa actitud hacia la esencia? Kripke sin duda tiene razón en que la idea de propiedad necesaria es mucho más que «una noción del filósofo sin ningún contenido intuitivo»[26], es bastante básica a nuestro modo de hablar. Pero, como muchas otras ideas parecidas (por ejemplo, «malo a secas y sin más, intrínsecamente malo», como en «el sexo prematrimonial es malo sin más»), se vuelve controvertida cuando alguien pregunta «¿cómo lo sabes?» Desde Kant, el espíritu de la filosofía ha sido verificacionista y dice que, si esta última pregunta no pudiera responderse, entonces la idea en cuestión también podría abandonarse. Por decirlo de otro modo, todo el espíritu de la filosofía desde Kant (y, de hecho, desde Descartes) ha sido no reconocer el tipo de distinción entre conceptos «epistemológicos» y «metafísicos» que Kripke quiere trazar.[27] Por tanto, lo que ha parecido natural hacer con el contenido intuitivo de la idea de «necesidad» es relativizarlo: decir que, en cualquier estadio dado de la investigación (o, si se prefiere, de la historia del lenguaje, o de la historia de la imaginación), puedes abocetar las cosas en esencias y accidentes al sano y viejo modo aristotélico —aprovechando, por ejemplo, la idea científica más reciente sobre de qué están hechas las cosas—. Desde esta perspectiva, el sentido común siempre será aristotélico, solo que lo será de una manera diferente en cada generación. No será un trasfondo de la investigación científica; será simplemente investigación congelada en algún punto que se elige. Podemos ser aristotélicos cuando nos las habemos con las necesidades del momento, pero deweyanos cuando contemplamos el tiempo y la eternidad.

Confieso que no tengo ni idea de cómo argumentar si ese espíritu verificacionista (del que los enfoques de Searle y de Putnam sobre los nombres y los conceptos son la floración más reciente) es un error a la escala de la historia del mundo o constituye un genuino progreso. Me inclino a decir que no hay delineado un verdadero problema. Así, por ejemplo, Kripke toma la cita de Searle de que «es un hecho necesario que Aristóteles posee la suma lógica, la disyunción incluyente, de propiedades que comúnmente se le atribuyen», y comenta que «semejante afirmación, si "necesario" está usado como lo he venido usando yo en esta conferencia, debe ser claramente falsa».[28] Desde luego que debe serlo, pero ¿por qué no tener simplemente dos sentidos de «necesario», uno para quien quiere que su referencia esté determinada, por ejemplo, por las raíces causales del uso del vocablo (como lo que dice Donnellan), y otro para quien no se preocupa de raíces y quiere hablar de quienquiera que fuera el hombre que…? ¿Tiene que ser uno de esos sentidos el correcto? Por decirlo de otro modo, ¿tiene que haber una noción de «referencia» tal que el segundo individuo se está refiriendo, lo quiera o no, al final de esa cadena causal? Bueno, esto solo empuja las cosas un paso más atrás. ¿Hay una relación llamada «referencia» que es algo que va más allá de —que es más metafísico que— averiguar de qué está hablando la gente a base de preguntarles (o de qué estás hablando tú a base de preguntarte a ti mismo)? Si semejante relación existe, ¿su existencia cambia en algo otros tópicos filosóficos?

He estado arguyendo que las nociones aristotélicas remozadas que esgrime Kripke no afectan al problema de la identidad mente-cuerpo, pero no estoy seguro de que no afecten a otros problemas filosóficos interesantes. Además, creo que Kripke tiene razón en intentar sistematizar algunas de las intuiciones de sentido común que han quedado perdidas en el reciente movimiento a relativizar todo lo habido y por haber. Tiene que existir un modo de que podamos adoptar la elevada perspectiva de historia del mundo que comparten Hegel, Dewey y Quine, y a la vez seguir hablando con Aristóteles y el vulgo, evitando así paradojas relativistas chillonas.

Nadie ha logrado tener tanto tacto hasta ahora, y a lo mejor algunas de las viejas distinciones aristotélicas ayudan. (Aunque, si lo hacen, su uso tendrá que disociarse de novísimas paradojas como la de que «no se puede describir propiamente una situación contrafáctica en la que hubiera habido unicornios»).[29] Pero antes tendremos que formarnos una idea más clara del método de la metafísica-abstraída-de-epistemología y sobre cómo dirimir desacuerdos en ese terreno.

Materialismo. Terminaré con unas pocas palabras sobre el Dolor y el ser doloroso. Con el primero me referiré a un tipo natural de particular (una sustancia inextensa) que aparece aquí y allá en los organismos sentientes y que algunos han sugerido que realmente *no* es inextenso, sino que tiene como naturaleza metafísica intrínseca la estimulación de fibras C. Lo segundo es una cualidad fenomenológica inmediata que los dolores normalmente tienen en los humanos. Algunos filósofos piensan que es una contradicción pensar en dolores no dolorosos. Dicen que el ser doloroso es de la esencia de los dolores. Afirman saber eso a priori —y tener una creencia para la que la evidencia empírica es irrelevante— en virtud de un análisis de significados, o por una puesta entre paréntesis fenomenológica, o por puro *nous* sin más. Normalmente no afirmarían saber a priori sobre la esencia de ningún tipo de particular que no sean estados mentales, pero es que estos, creen ellos, tienen esencias especialmente fáciles de conocer. Un filósofo que adopte esta posición pisa terreno firme para resistirse a la identidad mente-cuerpo. Lo único que falta es que desarrolle una epistemología y una metodología que dé cuenta de nuestro acceso a priori a lo mental. Eso no es fácil.

Los filósofos que se resisten a la identidad mente-cuerpo, pero que (quizá asustados por Wittgenstein) no están seguros de querer resucitar la visión cartesiana de que «nada es más fácil de conocer para la mente que ella misma», tratan de sortear la cuestión epistemológica. Dicen que lo que está en cuestión es la cualidad fenomenológica inmediata misma y que esta es «irreducible», que no puede identificarse con estimulaciones de fibras C, ni con la cualidad de tener uno estimuladas sus fibras C, ni con nada. Así que

podemos olvidarnos del Dolor y contemplar sin más esa cualidad irreducible. ¿De qué es cualidad? No de los dolores, ya no están. ¿De la experiencia? Desde luego, pero eso no ayuda mucho, ya que toda «cualidad inmediata (es decir, de la que se puede informar no-inferencialmente)» es una cualidad de la experiencia. ¿Qué decir de la experiencia? ¿Es el nombre de un haz de cualidades o es una sustancia inextensa? Si lo primero, ¿a qué va adosado ese haz de cualidades? Tal vez a cerebros; no parece que nadie tenga ninguna idea mejor. Si lo segundo, ¿cómo conocemos la esencia de esa sustancia inextensa? ¿Realmente estamos seguros de que es inextensa? ¿Qué método a priori aplicamos aquí para confirmar nuestra intuición? ¡Vuelta a la epistemología otra vez!

He fabricado estos dilemas para insinuar que preguntas como «¿sustancia o cualidad?», «¿esencia o accidente?», están fuera de lugar. Coincido con la idea wittgensteiniana (citada más arriba) de que por desgracia los filósofos han insistido en tomar las sensaciones como si fueran poco menos que objetos físicos y en aplicarles el mismo vocabulario que a los gatos, los demonios y las montañas. La razón por la que no deberíamos hacer eso no es, creo, que sus «lógicas» sean diferentes ni nada por el estilo, sino sencillamente que, si piensas en ellas de esa manera, surgen preguntas que por principio no tienen respuesta. Esas preguntas solían llamarse pseudoproblemas, y se decía que su irrealidad se revelaba una vez comprendíamos que todas las partes en la discusión coincidían en «los hechos» y que todo lo que quedaba era la cuestión «verbal» de cómo expresarlos. Hoy en día no estamos tan seguros de esta distinción fáctico-versus-verbal, pero creo que su fuerza está justamente en que ciertos desacuerdos son estériles en el sentido de que no engranan de ningún modo interesante con desacuerdos sobre cualquier otra cosa. («Diferencias que no suponen ninguna diferencia», en expresión de James.) Los materialistas discrepan entre sí y con sus oponentes sobre (a) si hay particulares llamados dolores correlacionados constantemente con fibras C estimuladas; o si más bien (b) esas fibras, cuando son estimuladas, tienen ciertas cualidades fenomenológicas horrorosas; o si más bien (c) esas fibras no tienen esas cualidades, sino que las

producen en alguna otra parte (en nuestra mente, digamos); o si quizás (d) «estimulación de fibras C» y «esa cualidad fenomenológica horrorosa» son simplemente dos nombres de la misma propiedad; o si más bien…[30] Pero la mayoría de la gente está de acuerdo en que, si la fisiología está a la altura de nuestras expectativas, un lenguaje de estados-cerebrales y uno de sensaciones servirán igual de bien para lidiar con los peligros de la vida. Admitir esto último le quita la mayor parte de su interés al dualismo, y una incondicional negativa austiniana a usar con las sensaciones el vocabulario «esencia-accidente, sustancia-cualidad» le quitaría al materialismo la mayor parte de su pegada. Una vez que nos libráramos del espectro de las sustancias-fantasma y las propiedades-fantasma, incluso si la fisiología no saliera bien (es decir, si no aparecieran correlaciones psicofísicas potables), nadie pensaría de todas formas que hay un problema mente-cuerpo.

En otro lugar he argüido que, una vez logremos poner en orden las cuestiones *epistemológicas* relativas a nuestro conocimiento de lo mental, no van a quedar problemas metafísicos en ese terreno (exactamente igual que no los había para Aristóteles).[31] Una razón por la que creo que es importante *no* recuperar nociones metafísicas aristotélicas (por vía de la filosofía del lenguaje) con relación al problema mente-cuerpo es que pienso que el problema lo creó una mala imagen del conocimiento humano y se disuelve con una mejor. Por decirlo sin ambages, la epistemología puede curar los males que causó la epistemología. Pero esta es una opinión que no puedo defender aquí. Espero, no obstante, que esta discusión de las ideas de Kripke sobre la identidad mente-cuerpo muestre que las objeciones intuitivas que suscita la afirmación de tal identidad ni ganan ni pierden fuerza si uno adopta las ideas sobre la referencia, el nombrar y la necesidad que Kripke propone.

---

[1] Me apoyo aquí principalmente en la discusión que presenta Kripke en «Naming and Necessity», en *Semantics of Natural Language*, ed. por Donald Davidson y Gilbert Harman. Boston: D. Reidel, 1972, pp. 253-355 [hay trad.

esp.: *El nombrar y la necesidad*, México, UNAM/Instituto de Investigaciones Filosóficas, 2005].

[2] *Ibid*., p. 331.

[3] *Ibid*., p.269.

[4] *Ibid*., p. 337.

[5] Ver «Naming and Necessity», p. 321, a propósito de gatos y demonios. No estoy sugiriendo que Kripke tenga ninguna simpatía por la variante de la «desaparición» de la teoría de la identidad. No es tan fácil decir que las cosas que pensábamos que eran sensaciones resultaron ser «sensaciones para los tontos», como decir que todas las cosas que pensábamos que eran gatos (y que en realidad eran demonios) resultaron ser «gatos para los tontos». Si uno cree, como Kripke, que una cualidad fenomenológica inmediata es una propiedad esencial de una sensación, no está claro que *pudiera* haber cosas como «sensaciones para los tontos», porque nadie podría ser burlado.

[6] Kripke, «Naming and Necessity», p. 326.

[7] *Ibid*., p. 340.

[8] *Ibid*., p. 339.

[9] *Ibid*., p. 324.

[10] Cf. *ibid*., p. 340.

[11] *Ibid*., p. 335.

[12] Piénsese, por ejemplo, en las explicaciones de Wolfe May y David Harrah sobre la influencia de los *Principia Mathematica* en *Proceso y realidad*, así como en su obvia influencia sobre el *Tractatus*. También en las campañas de Bradley y Peirce contra el tratamiento aristotélico estándar de los juicios sujeto-predicado, en las polémicas de Dewey y Quine contra la noción de «esencia», etc. etc. Para un tratamiento moderno de la doctrina aristotélica de que solo las sustancias tienen esencias (*Metafísica* VII, 4), véase Manley Thompson, «The Distinction between Thing and Property», en *The Return to Reason: Essays in Realistic Philosophy*, ed. por John Wild. Chicago: Henry

Regnery, 1953. Para una discusión de la opinión sugerida en *Metafísica* IV, 4 (1007a, 20 y ss.) de que la racionalidad y la necesidad sobreviven o sucumben según lo hagan las ideas de «sustancia y esencia», véase el estudio de Anscombe de ese pasaje en *Three Philosophers*. Ithaca, Nueva York: Cornell University Press, 1961.

13 Pienso aquí en el ataque de Wittgenstein en las *Investigaciones* a la idea cartesiana de «cualidad fenomenológica inmediata» y a las aristotélicas de «definición por género y especie» y «esencia», y su insinuación de que ambos casos ejemplifican la misma falsa imagen de la relación entre el lenguaje y el mundo.

14 Estas diversas tácticas wittgensteinianas o cuasi-wittgensteinianas se pueden encontrar en varios autores bajo diferentes formas. George Pitcher en *The Philosophy of Wittgenstein* (Englewood Cliffs, Nueva Jersey, Prentice-Hall, 1964) y John W. Cook en «Wittgenstein on Privacy» (*Philosophical Review* 74, n.º 3 [1965], pp. 281-314) afirman que la primera es la intención de Wittgenstein. La segunda representa el así llamado tratamiento «conductista» de Wittgenstein, refinado y modificado con fines materialistas por David M. Armstrong en *A Materialist Theory of the Mind* (Londres, Routledge, 1965) y por David Lewis en «An Argument for the Identity Theory» (*The Journal of Philosophy* 63, n.º 1 [1966], pp. 17-25). Yo he argumentado a favor de la tercera táctica en «In Defense of Eliminative Materialism» (*Review of Metaphysics* 24, n.º 1 [1970], pp. 121-121), contra la primera en «Wittgenstein, Privileged Access, and Incommunicability» (*American Philosophical Quarterly* 7, n.º 3 [1970], pp. 192-205), y contra la segunda en «Incorrigibility as the Mark of the Mental» (*Journal of Philosophy* 67, n.º 12 [1970], pp. 399-424). No obstante, no creo que la elección entre ellas sea importante en el presente contexto.

15 Véase, por ejemplo, David Pears, *Ludwig Wittgenstein*. Nueva York: Viking, 1970, pp. 100 y ss. y 159 y ss. [hay trad. esp.: *Wittgenstein*, Barcelona, Ed. Grijalbo, 1973]; y también Cook, «Wittgenstein on Privacy», pp. 313-314.

16 Kripke, «Naming and Necessity», p. 335.

17 *Ibid*.

18 Cabe pensar que el dilema que he presentado se funda en una mala comprensión de la idea de Kripke de los nombres propios. Pues esa idea quizá

parezca ser que siempre asociamos un tipo natural con un nombre propio desde la primera vez que se le aplica. Es decir, la idea de Kripke podría ser que «Héspero» no refiere a menos que Héspero sea una estrella; «Nixon», a menos que sea un hombre; etc. Pero esta no es de hecho la idea que él tiene:

> Algunas propiedades de un objeto pueden serle esenciales en cuanto que no habría podido dejar de tenerlas; pero esas propiedades no se usan para identificar el objeto en otro mundo posible, porque no hace falta semejante identificación; ni hace falta que las propiedades necesarias de un objeto sean las que se utilizan para identificarlo en este mundo, si de hecho se lo identifica en el mundo real mediante propiedades (p. 273).

Si Kripke aplicara este pasaje a tipos igual que a individuos —a términos como «dolor» y «gato» igual que a «Nixon» y «Héspero»—, entonces tendría que decir, o bien (a) que el dolor no necesita ser seleccionado mediante una propiedad necesaria (fenomenológica inmediata), sino que simplemente sucede que nos hemos fijado en una para seleccionarlo; o bien (b) que hay algo especial en los dolores (y, presumiblemente, en los estados mentales en general, quizá sea la «familiaridad directa») que hace que cacemos siempre su esencia —lo que *son*— a la primera. Pero entonces, una vez más, lo que estamos discutiendo es justamente si hemos cazado la esencia del dolor. El materialista piensa que no, y que tenga o no razón nada tiene que ver con sus opiniones sobre los conflictos de Kripke con Mill, de Kripke con Russell, etc.

[19] Cf. Kripke, «Naming and Necessity», pp. 329 y ss., donde se presenta la tesis.

[20] Cf. *ibid*., p. 322; pero aquí puedo estar malinterpretándole.

[21] *Ibid*., p. 327.

[22] *Ibid*., p. 328.

[23] *Ibid*., p. 322.

[24] Saul Kripke, «Identity and Necessity», en *Identity and Individuation*, ed. de Milton Karl Munitz. Nueva York, New York University Press, 1971, p. 145.

[25] Kripke, «Naming and Necessity», p. 321.

[26] *Ibid*., p. 265.

[27] *Ibid*., pp. 260-261.

[28] *Ibid*., p. 279.

[29] Saul Kripke, «Addenda to Saul A. Kripke's "Naming and Necessity"», en *Semantics of Natural Language*, ed. Por Donald Davidson y Gilbert Harman. Boston: D. Reidel, 1972), p. 763.

[30] La oración termina con esta elipsis en el original. —Eds.

[31] «Incorrigibility as a Mark of the Mental», citado más arriba en la nota 14, y «Cartesian Epistemology and Changes in Ontology», en *Contemporary American Philosophy*, ed. por John B. Smith. Nueva York, Humanities Press, 1970, pp. 273-292.

II
ESCRITOS POSTERIORES

CAPÍTULO 9

LA FILOSOFÍA COMO EPISTEMOLOGÍA: RESPUESTA A HACKING Y KIM

Tanto Ian Hacking como Jaegwon Kim hablan como si «conocimiento» y «filosofía» designaran tipos naturales; como si fueran la clase de cosa que tiene una esencia, una naturaleza que podemos captar bien o captar mal. Aunque los dos están dispuestos a prescindir del fundamentismo y de los «problemas eternos», piensan que hay un interés distintivo en filosofía y que tiene algo que ver con el conocimiento. Dice Hacking, por ejemplo, que «pensar sobre el conocimiento ha sido parte integral de la filosofía».[1] A Kim le gustaría que esta última fuera, como mínimo, «investigación *intra*paradigmática relativa a los aspectos conceptuales, de fundamento y regulativos, de un paradigma dado».[2] Para hacer más claro mi desacuerdo con estas afirmaciones, quiero distinguir dos sentidos relevantes de «filosofía» y de «conocimiento». El primer sentido en ambos casos es demasiado amplio para resultar interesante o para que sugiera la posibilidad de una indagación en esencias. El segundo es tan estrecho que solo nombra un utensilio gastado. Hacking y Kim trabajan con algo que, o bien representa un tercer sentido (invisible para mí), o bien es una oscilación entre el sentido estrecho y el ancho.

En su sentido ancho, «filosofía» significa, como ha dicho Wilfrid Sellars, «ver cómo las cosas, en el sentido más amplio del término, encajan entre sí, en el sentido más amplio del término».[3] En su sentido estrecho, puede definirse como un esfuerzo por mejorar la cantidad o la calidad de nuestro conocimiento a base de seleccionar un grupo privilegiado de aserciones, afirmar que están en cierta forma especial de correspondencia con algunas entidades extralingüísticas, e impartir luego esa relación especial a tantas aserciones más como podamos. En el sentido amplio, «filosofía» es

sinónimo más o menos de «alta cultura». En el estrecho, se usa para separar a pensadores como Platón, Locke y Dewey de pensadores como el Pseudo Dionisio, Goethe, Kierkegaard y Matthew Arnold.

En el sentido amplio de «conocimiento», el término se aplica a todas las aserciones garantizadas, o más bien, a aquellas que todavía parecerán garantizadas al final peirceano de una conversación habermasiana sin distorsiones prolongada indefinidamente. Entre las candidatas a esta última clase están, por ejemplo, «si la pintas de esa manera, parecerá contrahecha», «si lo edificas de ese modo, quedará feo», «cantidades iguales sumadas a cantidades iguales dan cantidades iguales», «debemos amarnos los unos a los otros», «Dios es Tres Personas y Una Sola Sustancia», «Wintergreen es el mejor candidato»[4], «*F* es igual a *m* por *a*» y «el gato está sobre el felpudo». En el sentido estrecho de «conocimiento», el término se aplica a una clase mucho más restringida cuyos límites exactos se deciden mediante una reflexión filosófica sobre cómo acortar la conversación por medios «metodológicos». Entre los candidatos a ser incluidos en esta segunda clase fueron populares «Dios es Tres Personas y Una Sola Sustancia» y «debemos amarnos los unos a los otros». Ahora mismo, los preferidos son «cantidades iguales sumadas a cantidades iguales dan cantidades iguales», «*F* es igual a *m* por *a*» y «el gato está sobre el felpudo». A grandes rasgos, una aserción es candidata a la clase estrecha, privilegiada, si no parece probable que nadie vaya a salir con algo nuevo que la haga parecer injustificada, o a sugerir un vocabulario nuevo que haga parecer obsoletos los términos en que se emitía.

Si «filosofía» y «conocimiento» se toman en el sentido estrecho, entonces es analítico que, como dice Hacking, «casi todos aquellos a los que incluimos en nuestro canon (de grandes filósofos) estuvieron fascinados por lo mejor del conocimiento de su época».[5] La razón de que pongamos a Platón y a Locke y a Dewey en la sección de filosofía de la biblioteca es que esperaban descubrir el secreto de algunas disciplinas que lo estaban haciendo muy bien en su tiempo y aplicar esa clave del éxito a nuevos campos. Ponemos al Pseudo

Dionisio y a Goethe y a Arnold en otra parte porque sus intentos de hacer que las cosas encajaran entre sí, de ver la vida sin desmayo y verla completa, no tenían mucho que ver con aislar aserciones privilegiadas y tratar de especificar exactamente por qué eran privilegiadas. Pero creo que merece la pena señalar que, cuando Hacking dice que si la metáfora de un Ojo Mental es tan popular es porque de los antiguos «la contribución más permanente al conocimiento es la geometría»,[6] está incurriendo en un montón de peticiones de principio. Desde luego que los antiguos a los que él llama «filósofos» hicieron mucha geometría. Pero ¿de verdad la obra de Euclides es más permanente que, digamos, el conocimiento de cómo gobernar ciudades mediante asambleas democráticas, el de cómo plasmar en piedra el cuerpo humano en movimiento, el de cómo secularizar, intelectualizar y sublimar nuestros instintos religiosos? Para mí, que Hacking elija sin titubear la geometría es un ejemplo de la costumbre de los filósofos de usar el sentido estrecho de «filosofía» para demarcar algo que luego puedan hacer pasar por un tipo natural. Las preguntas «¿y si…?» que plantea Nietzsche en *El nacimiento de la tragedia* y Heidegger en *Introducción a la metafísica* me parecen modos útiles de recordarnos a nosotros mismos que había muchos caminos que Occidente podía haber seguido en el siglo IV a. C., y que eran solo unos pocos antiguos elegidos a dedo por nosotros los que «sabían bien» que «su contribución más permanente al conocimiento» era la geometría.

Volviendo ahora a la distinción entre el sentido amplio y el estrecho de «filosofía», mi objeción a la filosofía en su sentido estrecho —el de que se centra en algo como la epistemología— es histórica y empírica más que de principio. A mí me parece una empresa que sencillamente no dio resultado. Seguramente no merece la pena esforzarse en averiguar cuáles de nuestras aserciones garantizadas seguirán haciéndose al final de la investigación. Incluso si acertamos en qué contribuciones nuestras serán duraderas, no parece probable que vayamos a encontrar en ellas algún ingrediente especial que podamos aislar y utilizarlo para hacer más contribuciones. Ahora bien, no creo que haya modo de

demostrar limpiamente a priori que es *imposible*. Lo que digo es que la historia de los intentos de hacerlo es lo bastante larga y lo bastante decepcionante como para hacernos pensar que de entrada no era un proyecto muy bueno. Pero sin este proyecto en mente no veo que podamos darle sentido a eso que Hacking llama «intentar entender la posibilidad y la naturaleza de los diversos tipos de conocimiento y de estilos de razonamiento».[7]

En una interpretación, la frase de Hacking puede tomarse simplemente como una referencia a la filosofía en el sentido amplio, algo que hacía Goethe no menos que Dewey. Lo que intentaríamos sería ver en qué se parecen y en qué difieren las diversas actividades humanas que culminan en aserciones garantizadas. Eso es lo que harían los filósofos si adoptaran lo que yo llamo una actitud hermenéutica, más que epistemológica, hacia la cultura; actitud que llevaría, sospecho, a una pérdida gradual de su sentido de autonomía profesional y a la paulatina fusión de la filosofía con la historia y la literatura. Pero doy por sentado que Hacking se opondría a semejante difuminación de géneros. A él no le parece atractivo el papel que yo recomiendo para el filósofo de ser «un diletante informado, que concilia o trasciende disciplinas y discursos». En concreto, piensa que «la historia y la sociología no son sucesoras de la epistemología, sino parte de las herramientas con las que el epistemólogo ahondará en sus preguntas».[8] Mi problema es que no veo qué preguntas podrían ser esas.

La frase de Hacking me parece tanto más desconcertante por cuanto él y yo somos admiradores de Foucault. Yo tomaría *Las palabras y las cosas* (o *Vigilar y castigar*, o *Historia de la locura en la época clásica*) de Foucault como un buen ejemplo de cómo la historia y la sociología de diversas disciplinas pueden hacer más por nosotros que la epistemología. Sé que Hacking entiende que Foucault está haciendo algo distintivamente epistemológico, pero a mí me cuesta ver qué es. Ni me ayudan tampoco sus referencias a Chomsky y a Kuhn. No veo que Chomsky esté diciéndonos nada constructivo sobre la naturaleza y la posibilidad del conocimiento humano, sino apuntando una tesis negativa sobre los programas de

investigación tipo Skinner. De la misma manera, no veo que Kuhn nos esté diciendo nada de esa índole, sino apuntando una tesis negativa sobre la llamada «visión heredada» en filosofía de la ciencia. Ambas tesis negativas han sido extraordinariamente útiles, pero soy incapaz de ver que puedan tomarse como respuestas a preguntas epistemológicas. En mi opinión, las autodescripciones que hace Chomsky de su propio trabajo en términos de racionalismo e innatismo, y el intento de Wolfgang Stegmüller de utilizar a Kuhn para reeditar los criterios de elección teórica al viejo estilo,[9] son exactamente lo que no hay que hacer. Son intentos de verter vino nuevo en odres viejos, matando así el sabor. Sospecho que el tratamiento que hace Hacking de Foucault es algo parecido, pero puede ser solo que no esté entendiendo correctamente su visión de Foucault.

Para cerrar mi réplica a Hacking, puedo resumir lo dicho de la siguiente manera: solo tiene sentido preguntar por la naturaleza de algo si tienes alguna perplejidad concreta que quieres resolver. En el caso de «conocimiento» en el sentido *amplio*, no creo que tengamos una idea clara de qué queremos decir con «la naturaleza del conocimiento», porque no podemos señalar con el dedo una dificultad en particular que «el conocimiento de la naturaleza del conocimiento» resolvería. En abstracto, no parece que las posibilidades de dar con algo que haga garantizadas a todas las aserciones garantizadas sean mayores que las de dar con algo que haga buenos a todos los buenos modales. Si lo que está en cuestión es el conocimiento en sentido *estrecho* —si estamos preguntando qué pueden tener en común los axiomas de Euclides, las leyes de la mecánica y los informes sobre la localización de los gatos—, entonces doy por hecho que Hacking, Kim y yo coincidimos en que en eso consiste exactamente la desacreditada búsqueda de fundamentos. De modo que me inclino a sospechar que «la naturaleza del conocimiento» es una expresión vacía. «La naturaleza de la ciencia» quizá parezca un poco más prometedora, pero dudo que lo sea. No creo que vayamos a encontrar una estrategia común que fuera utilizada para poner sucesivamente «sobre bases

científicas» las matemáticas, la física, la química, la biología, la economía y la filosofía. En todos los intentos a los que hemos asistido de convertir en científico esto o aquello, la intromisión metodológica nunca fue buena cosa. Lo que obró el milagro fue que apareció alguien con un vocabulario sustancialmente nuevo que funcionó. Creo que la sabiduría en este terreno empieza con ver como científicos a Galileo y a Dalton y a Darwin porque tuvieron éxito, no que tuvieron éxito porque fueron científicos. A mí me parece que «la senda segura de una ciencia» no es más que el galardón que se otorga a cualquier disciplina que parezca haber hecho algo muy impresionante y ahora es el escenario de una razonable cantidad de consenso y de trabajo en equipo. Para mí, «ciencia» es una noción sociológica más que epistemológica.

En cuanto a la posibilidad del conocimiento, la pregunta solo me parece interesante si alguien puede sugerir por qué no habría de ser posible. Desde Kant, los filósofos han adquirido la mala costumbre de explicar lo que hacen diciendo que están investigando las «condiciones de posibilidad» de esto y aquello, en algún sentido de «condición» que es tan vago como el de Kant de *Grund*. No obstante, quizá Hacking tenga en mente condiciones claramente causales como las que describe Foucault: cambios en la jerga y en las instituciones que dejan un hueco para que algunos genios se hagan un nombre. Pero no veo cómo llegar desde el viejo sentido kantiano de «posibilidad» a este sentido sociológico. A Hacking no le gusta mi constante referencia a «prácticas» en *La filosofía y el espejo de la naturaleza*. Dice él: «pero hablar de prácticas no es ni tan siquiera empezar a preguntar algo».[10] A mí me parece que esa observación invita inmediatamente a preguntar qué preguntas hacer. Mi razón para hablar tanto de prácticas es la misma que la de Wittgenstein: hace tanto más difícil repetir las malas preguntas de siempre.

Dice Hacking que me equivoco al afirmar que, «cuando una práctica ha continuado el tiempo suficiente, las convenciones que la hacen posible son relativamente fáciles de aislar». «Porque», afirma él, «no podemos aislar las convenciones pertinentes» que «han

hecho posible la práctica de la química desde los tiempos de Lavoisier».[11] Pero a lo que yo me refería con convenciones era simplemente a qué aprende una persona al aprender a ser un químico, por ejemplo. Desde luego, no le damos una lista de convenciones para que la memorice, como le damos las Ordenanzas Militares a un soldado. Pero el estar en la senda de una ciencia lleva consigo el que haya un procedimiento de iniciación —por ejemplo, la educación universitaria— que haga de ti un científico. Lo que aprendes es cómo llevar a cabo una práctica, y «familiaridad con las convenciones relevantes» parece un nombre razonable, si no particularmente informativo, para una parte de lo que comporta eso. Hacking piensa que es demasiado poco informativo. Así que mi diferencia básica con él es si podemos esperar más información. Yo creo que no, pero, repito, lo único que tengo para justificar mi pesimismo es el registro de anteriores fracasos. Llegado un cierto punto, es sensato tomar las decepciones pasadas como una razón para abandonar un proyecto que todavía puede parecer bueno sobre el papel. La historia de las respuestas a las preguntas sobre la naturaleza y la posibilidad del conocimiento me parece demostrar que algo había mal en el proyecto epistemológico.

* * *

Paso ahora a la observación de Kim de que no se puede culpar al platonismo de haber engendrado la epistemología. Kim considera que nociones como «correspondencia con la naturaleza» y «representación exacta», tal como Platón las usa, son inofensivas y de sentido común, a la vez que se suma a mis dudas sobre el mentalismo cartesiano y el universalismo kantiano.[12] Tengo que rechazar su oferta de conciliación, pues son precisamente esas nociones platónicas las que le permiten utilizar la distinción «cognitivo vs. no-cognitivo» para criticar mi concepción de la filosofía. Esta última distinción me parece ininteligible sin el supuesto platónico de que algunas de nuestras aserciones garantizadas se corresponden con la realidad y otras no. Sin esa idea nunca habríamos desarrollado la filosofía en el sentido estrecho ni le

habríamos dado al término «conocimiento» su significado filosófico distintivo. No habríamos pensado que «2 + 2 es 4» y «el gato está sobre el felpudo» tienen una clase de verdad más clara y más básica que «debe haber más amor en el mundo». Sin las nociones de «correspondencia» y «representación exacta», no habríamos tenido ese proyecto en el que la filosofía en sentido estrecho se ha metido de buscar lazos entre los pensamientos (o las oraciones) y el mundo en derredor.

No obstante, podemos abandonar las nociones de correspondencia y representación y seguir teniendo lo que Kim llama «la función asertórica del lenguaje». Para mí, la gloria de Dewey, Wittgenstein y Sellars es justamente haber mostrado cómo podemos distinguir las aserciones de otros ruidos (por ejemplo, las toses y las órdenes) desde un análisis de «juegos de lenguaje», sin emplear nociones platónicas, no-sellarsianas.[13] Creo que la virtud de un análisis así —y, más en general, de este tipo de aproximación al pensamiento y al lenguaje— es precisamente que elimina en parte la tentación de pensar que «debe haber más amor en el mundo» posee una suerte de verdad no perspicua, dudosa, de segunda clase, comparada con la inteligible y clara de «el gato está sobre el felpudo». Los análisis de juegos de lenguaje nos permiten decir que ambas oraciones son verdaderas sin tener que preocuparnos de qué las «hace» verdaderas.

Sospecho que, para Kim, esa es una preocupación que debemos tener, y que, como buen platónico, piensa que la verdad sin hacedores-de-verdad no es lo bastante buena. No sé cómo debatir ese punto, salvo diciendo que aquí de nuevo tenemos que apoyarnos primordialmente en cómo ha rendido, por así decir, el programa de investigación platónico. No creo que analizar lo que queremos decir con «verdadero» vaya a servir de mucho, porque el platónico y el wittgensteiniano tienen ideas bien distintas de lo que es «dar un análisis». Sospecho que la posición de uno sobre este asunto se va a decantar en última instancia echando una mirada al panorama filosófico contemporáneo y decidiendo por dónde tirar, dónde se está haciendo un trabajo más prometedor. La cuestión de si atenerse o

no al platonismo —a lo que Heidegger llamaba «la metafísica de la presencia»— es la que en la actualidad divide a filósofos «analíticos» y «continentales». Gente como Derrida y Foucault dan por sentado el punto de vista anti-platónico, se han acostumbrado a una verdad sin hacedores-de-verdad y han intentado concebir una forma de vida intelectual que no gire en torno a la distinción cognitivo vs. no-cognitivo. A mí esa forma de vida me parece que promete más que la que defendían y explicaban Russell y Reichenbach. Pero, repito, no tengo argumentos contundentes que ofrecer, solo propongo que lo veamos como una elección práctica más que teórica, más como una elección sobre qué disciplina practicar que como una elección entre dos teorías en pugna dentro de una misma disciplina.

De todos modos, se sitúe uno donde se sitúe en relación con el hacer-verdadero y con la metafísica de la presencia, se debería tener clara la diferencia entre decir

(a) no hay nada así como hacedores-de-verdad y
(b) la aprobación social es el único hacedor-de-verdad

Yo estoy diciendo lo primero más que lo segundo. Kim habla como si yo «socializara» la verdad, pero sería más exacto decir que, como Dewey y James, intento moralizarla. La aprobación social presente —el estar garantizada ahora— desde luego que no asegura que una aserción seguirá estando garantizada al final de la investigación, que sobrevivirá a las vueltas y revueltas conversacionales del futuro. Ni tampoco una sociedad va a estar nunca en situación de declarar que la conversación ha terminado. Es cierto que cualquier sociedad puede intentar declarar eso —y la mayoría lo han hecho— respecto de ciertos temas de conversación. Por consiguiente, han «distorsionado la comunicación», por usar la expresión de Habermas, han restringido la conversación. Pero no deberían haberlo hecho. La razón no es que una comunicación no distorsionada sea el mejor modo de conseguir representaciones exactas de la realidad, sino que la comunicación no distorsionada es algo moralmente bueno. Mientras que Hacking cree que lo que hace occidental a Occidente es su adhesión al «conocimiento de la verdad

como un valor por sí mismo»,[14] yo creo que lo que nos hace occidentales es nuestra disposición a seguir hablando, a probar nuevos vocabularios, a romper la convención con la esperanza de una conversación todavía más interesante. Hacking y Kim tienden a identificar la virtud especial de Occidente con un interés en captar correctamente las cosas. Yo la identifico con dejar libre a la gente para hablar de formas en que nunca hablaron antes, la identifico con Pericles más que con Euclides. Mientras pensemos que la verdad la hacen poderes distintos a nosotros, buscaremos restricción y seguridad —la senda segura de una ciencia— en vez de poner nuestra libertad a prueba. La alternativa no es pensar que la verdad la hacemos nosotros, sino que no se hace en absoluto. Si, como James, vemos lo verdadero como lo que es bueno en el ámbito de la creencia, entonces rechazaremos la pregunta «¿qué hace verdaderas a las verdades?» por la misma buena razón por la que rechazamos la pregunta platónica «¿qué hace buenos a los bienes?» Ambas son preguntas que, cuanto más intentamos contestarlas, peores parecen.

Pasando ahora de lo que Kim dice sobre el conocimiento y la verdad a lo que dice sobre la filosofía, quisiera hacer algunas críticas a su modo de verla como Sirvienta de la Ciencia. A grandes rasgos, mi idea es que la ciencia no necesita lo que Kim llama una «investigación intraparadigmática relativa a los aspectos conceptuales, de fundamento y regulativos, de un paradigma dado».[15] Puede que sea por mi ignorancia de la historia de la ciencia, pero no se me ocurre ningún buen ejemplo de una disciplina a la que meditar sobre los aspectos de fundamento, conceptuales y regulativos —a la que volverse filosófica— le haya hecho mucho bien. Me parece que esa clase de meditación es un fenómeno tipo Lechuza de Minerva, causado por lo que Kuhn llama la acumulación de anomalías que precede a una revolución, o bien por puro aburrimiento. Volviendo a algo que ya dije antes, lo que hace que la conversación continúe es alguien con una idea brillante sobre un nuevo modo de hablar. La autorreflexión solo me parece útil cuando un paradigma antiguo, maduro, satisfecho de sí mismo, ha sido

desafiado descaradamente por una de esas nuevas ideas. Es una maniobra defensiva, que solo tiene utilidad cuando efectivamente hay un enemigo del que defenderse. Una práctica solo necesita molestarse en hacerse «internamente coherente y consistente»,[16] creo yo, si existe una duda concreta sobre su bondad. Cuando no hay tales dudas, aparecen los típicos libros estériles sobre «Fundamentos conceptuales de…». A mí me parece mejor interpretar el imperativo filosófico, socrático, de ser críticos y reflexivos como el de tratar de tener ideas nuevas, no el de examinar «cuestiones conceptuales, de fundamento y regulativas» en abstracto. Entiendo que el valor de Occidente es hacer que la conversación siga abierta y moviéndose, no asegurarse de que la investigación tiene bases sólidas o está en contacto con lo real. Así que no creo que la sistematicidad sea un valor intrínseco. Es simplemente el estilo normal de razonamiento intraparadigmático.

La alusión a la filosofía sistemática me lleva a lo último que quiero decir sobre el artículo de Kim. Afirma que, en mi concepción, la filosofía no apunta «a la obtención de la verdad intelectual»;[17] se vuelve no cognitiva, algo así como una forma de arte. Aquí insisto otra vez en que apunta exactamente a la clase de verdad a la que apuntan la física y la política. En la medida en que una actividad humana produce aserciones, apunta a la «verdad intelectual», la única que hay. Al distinguir entre filosofía «sistemática» y «edificante»[18], quizá he dado la impresión de estar invocando una forma de la distinción cognitivo vs. no-cognitivo, pero es una impresión equivocada. Más bien quiero difuminar la distinción entre arte y ciencia diciendo que la interesante es la que se da entre formas normales y anormales de uno y otra.

Creo que he causado una confusión innecesaria con mi uso de la palabra «edificante», ya que Hacking considera que aplicarla a Wittgenstein, por ejemplo, es hacerle de menos. Además, varios reseñadores de *La filosofía y el espejo de la naturaleza* han supuesto que «filosofía edificante» y «hermenéutica» eran más o menos sinónimos. Mi intención era mantenerlas bien separadas. La filosofía hermenéutica es justamente lo que Hacking llama «un interés por

los estilos de pensamiento» cuando ese interés se despreocupa de «la naturaleza y la posibilidad del conocimiento» y de la distinción cognitivo vs. no-cognitivo. Cualquiera puede hacer ese tipo de filosofía simplemente comentando el estado de la cultura, la escena pasajera, desde una cierta actitud pragmática. Mi libro es filosofía hermenéutica en ese sentido, un sentido en el cual la filosofía se mezcla con la historia y la sociología. Por el contrario, «filosofía edificante» es algo que solo un ocasional genio puede darnos. Pocas veces por siglo tenemos un filósofo que se adentra en un modo totalmente nuevo de hablar de las cosas en general. Wittgenstein, Heidegger y Dewey me parecen los candidatos de nuestro siglo. Ellos no intentaron ser críticos y reflexivos desde dentro de un paradigma maduro y comprehensivo, sino que desafiaron el vocabulario imperante de la alta cultura con ideas mejores.

Yo no sé cómo imitar a esos filósofos edificantes, y desde luego no estoy sugiriendo que la filosofía como disciplina académica deba dejar de ser sistemática y se vuelva edificante. Mientras siga siendo una disciplina, por supuesto que será sistemática. No hay modo de formar a los estudiantes para que sean edificantes, no más que de formarles para ser originales. Podemos formarles para ser hermenéuticos dando Nietzsche en vez de Russell, Derrida en vez de Trilling, Foucault en vez de Parsons. Pero eso no producirá necesariamente resultados edificantes. De hecho, abre la aterradora posibilidad de tesis doctorales sobre «los aspectos conceptuales, de fundamento y regulativos» de la hermenéutica.

* * *

Me gustaría concluir con algunas palabras a propósito de dos de las afirmaciones de Hacking en que sostiene que yo no describo bien la historia. Dice que Kant no fue el primero en usar la distinción entre sentido e intelecto, entre particular y universal, como hilo conductor para escribir la historia de la filosofía. En eso tiene razón. Exageré el punto cuando dije que Kant nos dio una «historia de nuestra materia». Mi punto era que, hasta los tiempos de Kant, había muchas distinciones (por ejemplo, fe y razón, naturaleza y

espíritu) que podrían haberse usado para escoger un canon, para dividir la filosofía en sentido amplio entre filosofía en sentido estrecho más un montón de pensadores no filosóficos. Kant suministró el paradigma que hizo que sus distinciones fueran las que había que buscar. El nuevo tipo de «historias de la filosofía» que se escribieron en las décadas de 1780 y 1790 me parecen buena evidencia de esto. Sin duda Hacking tiene razón en que Kant tomó ese hilo de Leibniz, quien a su vez lo tomó de la querella entre escolásticos agustinianos y aristotélicos sobre si había algo en el intelecto que no estuviera antes en los sentidos. Traer a primer plano esa distinción quizá se remonte hasta la «lucha entre los dioses y los gigantes» de Platón.

En cambio, no está claro que los trozos de Platón y Aristóteles que Leibniz y Kant consideraron centrales fueran centrales para ellos mismos. Hay muy poca evidencia de que lo fueran. Una vez que tenemos el canon que nos da la historiografía kantiana, no debería sorprendernos ver aparecer distinciones kantianas por todas partes y que se destaquen los textos donde se mencionan esas distinciones. Pero no creo que tengamos una buena respuesta para la pregunta «¿es ese canon el "correcto", como sugiere Hacking, o no es nada más que el "neo-kantiano", como estoy sugiriendo yo?» Ni siquiera estoy seguro de qué clase de pregunta es esa. Es un poco como la de «¿los poetas que Hume enumera en "Del criterio del gusto" son la piedra angular del canon correcto, o solo un canon dieciochesco típico?» No estoy seguro de qué sería «describir bien la historia» en el caso de semejantes preguntas. Sí creo que hay evidencia de que Kant fue la figura que permitió a quienes ostentaban las antiguas cátedras que Hacking menciona desarrollar la autoimagen que hoy tenemos nosotros: la de practicantes de una disciplina «formal» que es autónoma y crítica, que ni depende del estado actual de la ciencia ni responde de los sentimientos religiosos de la comunidad. Esa no era la imagen que tenían los ocupantes pre-kantianos de esas mismas cátedras, y creo que necesitamos una explicación de cómo la obtuvimos, de la «fractura» foucaultiana que comporta. Lo que Hacking llama «la lectura estándar» no ayuda a eso (aunque *Las palabras y las cosas* de Foucault sí).

Otra idea que defiende Hacking en la primera sección de su artículo, comentando la p. 149[19] de *La filosofía y el espejo de la naturaleza*, lleva el encabezado: «la validez del interés en lo universal y lo particular». Ahí dice que mi proyecto no quiere aducir, como Russell, que la distinción particular-universal es un accidente del lenguaje, ni que es un error, como aduce Ramsey. Desde luego que no, y no veo por qué Hacking cree que debería. En el sentido en que esa distinción es un «accidente del lenguaje», todas lo son. No sé en qué consistiría una distinción que fuera «*esencial* al lenguaje» en cuanto que opuesta a una para cuya traducción por paráfrasis pudiera usarse cualquier lenguaje. Ni veo en qué sentido se la podría llamar «un error». Como mucho, Ramsey demostró que puedes pasarte sin ella para diversos propósitos. El debate entre Strawson y Quine en torno a la irreductibilidad de los términos singulares tiene su interés, pero el tema adolece de la misma vaguedad que todos los debates sobre «reducibilidad»: ¿reducibilidad con qué propósito? Hacking parece pensar que yo debería argüir si la distinción particular-universal se corresponde con la realidad o no, si está «realmente ahí». Decir si se corresponde o no es lo último que yo haría. Lo único que sí deseo decir es que, como la distinción entre fe y razón y entre naturaleza y espíritu, llevamos demasiado tiempo discutiéndola para lo poco que hemos sacado en limpio. No hay nada de «inválido» en interesarse por tales distinciones. Es solo que —como las distinciones entre arrianos y ortodoxos, católicos y protestantes, marxistas y no-marxistas— ya no parece que tengan mucho sentido. Quizá solo lo parezca. Como dice Kim, puede ser que nos hayan caído en suerte «las preguntas filosóficas correctas».[20] Pero tomaría eso como una predicción de que va a llegar un genio que revitalice los problemas de empirismo o racionalismo, nominalismo o anti-nominalismo, idealismo o realismo, de forma que ya no parezcan meramente escolásticos como parecen hoy. Puede ocurrir. Así que puede que la filosofía sistemática, epistemológicamente centrada, regrese con nuevos bríos. Pero, ahora mismo, la filosofía sistemática parece en bastante mala forma. No puedo demostrar que no vaya a reponerse, pero no creo que

debamos esforzarnos por mantenerla con vida solo porque no tenemos mejor idea de lo que hacer.

---

[1] Ian Hacking, «Is the End in Sight for Epistemology?», *The Journal of Philosophy* 77, n.º 10 (1980), p. 586.

[2] Jaegwon Kim, «Rorty on the Possibility of Philosophy», *The Journal of Philosophy* 77, n.º 10 (1980), p. 595.

[3] Wilfrid Sellars, *Science, Perception, and Reality*. Nueva York: Humanities Press, 1963, p. 1. —Eds. [Hay trad. esp.: *Ciencia, percepción y realidad*. Madrid, Tecnos, 1971.]

[4] N. del T.: John P. Wintergreen es el protagonista del musical *Of Thee I Sing*, de los hermanos Gershwin sobre un libreto de Kaufman y Ryskind. La obra es una parodia política, en la que Wintergreen concurre a la presidencia de Estados Unidos desde la «plataforma del amor».

[5] Hacking, «Is the End in Sight for Epistemology?», p. 585.

[6] *Ibid*., p. 584.

[7] *Ibid*., p. 586.

[8] *Ibid*., p. 587.

[9] Wolfgang Stegmüller, *The Structure and Dynamics of Theories*. Nueva York, Springer-Verlag, 1976. —Eds. [Hay trad. esp.: *Estructura y dinámica de teorías*. Barcelona, Ed. Ariel, 1983.]

[10] Hacking, «Is the End in Sight for Epistemology?», p. 587.

[11] *Ibid*.

[12] Kim, «Rorty on the Possibility of Philosophy», pp. 589-590.

[13] «Asserting», de Brandom, es un ejemplo particularmente elegante de lo que puede hacerse en esa línea. «Asserting», *Noús* 17, n.º 14 (1983), pp.

637-650.

[14] Hacking, «Is the End in Sight for Epistemology?», p. 585.

[15] Kim, «Rorty on the Possibility of Philosophy», p. 595.

[16] *Ibid*., p. 593.

[17] *Ibid*., p. 594.

[18] N. del T.: la traducción de «*edifying*» por «edificante» está consolidada en las versiones al castellano de Rorty desde la aparición de *La filosofía y el espejo de la naturaleza* y forma parte ya del léxico rortyano en nuestro idioma, por lo que parece aconsejable respetarla. Pero sí merece la pena observar que induce fácilmente a error, pues «edificar» tiene para nosotros connotaciones (no tan inequívocas en inglés) que van justo en la dirección contraria, la de «infundir sentimientos de piedad y virtud» o, más literalmente, «hacer o construir un edificio; establecer, fundar» (RAE); en suma, producir algo sólido e inculcar aquiescencia a sus normas, que es justamente lo que caracteriza para Rorty a la filosofía de signo opuesto, la «sistemática». Quizá «iluminadora» o «instructiva» habrían recogido mejor que «edificante» la idea de Rorty de un estilo de filosofía que se sitúa en la periferia de la norma y la desestabiliza.

[19] N. del T.: la referencia es a la primera edición en inglés del libro de Rorty.

[20] Kim, «Rorty on the Possibility of Philosophy», p. 595.

CAPÍTULO 10

LA EPISTEMOLOGÍA NATURALIZADA Y LAS NORMAS: RESPUESTAS A GOLDMAN Y FODOR

Todo el mundo es partidario de conocer cuanta verdad sea posible, del mismo modo que todo el mundo es partidario de hacer todo el bien que se pueda. Pero si cabe o no decir algo general y útil sobre cómo abordar cualquiera de las dos cosas, y si puede haber una teoría general interesante sobre el conocimiento, o la verdad, o la bondad, ahí las opiniones difieren. Igual que a Aristóteles no le convencían los intentos de Platón de definir términos como «virtuoso» y «piadoso» y «bueno» de un modo que pudiera ser útil para quienes no tuvieran seguro qué acción es la correcta o cómo es una vida buena, así tampoco a filósofos como Reid y Dewey les convencieron los esfuerzos de los filósofos post-cartesianos de conseguir teorías del conocimiento y de la verdad que les resultaran útiles a quienes buscan verdades. Los teóricos de la ética aristotélicos contemporáneos, como Iris Murdoch, recelan de los principios morales generales. Insinúan que, en caso de aprieto, siempre te dejan tirado y te las tienes que arreglar tú solo. Los autores contemporáneos sobre ciencia, como Paul Feyerabend y Marshall Swain, recelan del tipo de preceptos generales que propugnaba Descartes en el *Discurso del método* y en las *Reglas para la dirección del espíritu*. Por ejemplo, cuando dice que debemos «conducir por orden [nuestras] reflexiones comenzando por los objetos más simples y más fácilmente cognoscibles, para ascender poco a poco, gradualmente, hasta el conocimiento de los más complejos»,[1] eso parece de tanta ayuda como que te digan que averigües si puedes querer que la máxima de tu acción sea una ley universal para todos los agentes racionales. Hay demasiados criterios de simplicidad —demasiados vocabularios en los que términos distintos se toman como indefinidos y primitivos— como para que la

regla de Descartes sirva de mucho, igual que hay demasiados vocabularios en los que puede describirse una determinada acción como para que la de Kant sirva de mucho tampoco.

Simpatizo con la opinión de que perseguir reglas para buscar el conocimiento nunca ha conducido a ninguna parte, y que, por consiguiente, la de «epistemología» puede que haya sido una mala idea; como simpatizo con quienes piensan que los principios morales generales no han contribuido mucho a guiar la acción, y que la idea de «teoría ética» quizá no haya sido buena tampoco. Coincido con lo que comenta Marshall Swain a propósito de otro artículo de Alvin Goldman[2]: dice Swain que él es «escéptico sobre la probabilidad de que haya progreso filosófico [en epistemología normativa] … como [lo es] sobre los sistemas que pretendieran decirnos cómo llevar la vida buena».[3] No creo que haya ni pueda haber argumentos definitivos de tipo general en contra de la posibilidad misma de reglas epistémicas o principios morales útiles, pero sí que su búsqueda ha durado ya lo bastante como para considerar si compensa seguir intentándolo. Si uno hace caso a Murdoch y Sartre en ser escéptico respecto de la teoría ética, y a Feyerabend en serlo respecto de la idea de «método científico», dirá que lo que de verdad necesitamos son ejemplos y no reglas o principios. Desde esta perspectiva, para aprender a actuar correctamente y a vivir bien uno leerá a Plutarco o a Balzac, más que a Kant o a Mill. Pero el mejor medio de todos será asociarse con el tipo correcto de personas. Análogamente, el mejor medio para dirigir tu espíritu y progresar en la disciplina científica, o en cualquier otra disciplina intelectual que elijas, será estudiar logros paradigmáticos recientes en ella. Pero será todavía mejor iniciarse en lo que Kuhn llama una «matriz disciplinar»: adquirir el saber-hacer de aquellos que en la práctica están logrando progresos en el campo en cuestión. Ninguno de estos medios es infalible, pues uno puede involucrarse con un grupo moralmente malo o iniciarse en los arcanos de un programa de investigación científica que está a punto de ser desplazado por el trabajo de un brillante revolucionario. Pero un principio no te ayudará a evitar tales inconvenientes.

Vista la epistemología de este modo, lo importante del *Discurso del método* de Descartes fue que dirigió la atención de la gente hacia los logros de Galileo y la apartó de la tradición aristotélica. Eso fue bueno, pero Descartes por desgracia dio a entender que Galileo había descubierto «cómo ser científico», no que descubrió solo un vocabulario mejor que el de Aristóteles para describir el movimiento. Popularizó así la infeliz idea de que Galileo había sido «racional» de un modo en que Aristóteles no lo había sido, y de que existen algo así como «cánones del procedimiento científico» cuya obediencia podía conducir en otros campos a descubrimientos comparables a los de Galileo en mecánica o a los del propio Descartes en geometría. Como he argumentado en otro lugar,[4] es dudoso que tales ideas hubieran podido sobrevivir de no ser porque Locke aceptó el reto y desarrolló la noción de que la racionalidad podía estudiarse estudiando la mente como si fuera una máquina para producir verdad, máquina que recibía como input «ideas simples» y producía como output cosas como las leyes de Newton. Esta propuesta armonizaba con la idea de ramistas y leibnizianos de fabricar «análisis» por procedimientos cuasi-mecánicos. Con el tiempo, la combinación condujo al Principio I de Goldman:

> *El estado doxástico D (en el momento t) de una persona está justificado si y solo si D es el resultado de una historia de transiciones entre estados-cognitivos conformes con (permitidas por) las reglas epistémicas correctas.*[5]

Desde el punto de vista escéptico que he propuesto, la pregunta que suscita este principio es «¿hay reglas de ese tipo que enseñen algo más que lugares comunes como "intenta evitar inconsistencias" y "mantén los ojos abiertos a nueva evidencia"?» La sospecha de que no las hay es lo que lleva a gente como Gilbert Harman a arrugar la nariz ante la idea de que las reglas de la lógica deductiva puedan conducirle a uno desde «enunciados básicos» a nuevos descubrimientos interesantes.[6] La vieja noción de que ser racional es arrancar desde el tipo correcto de premisas y luego «pensar lógicamente», común a Aristóteles y Locke, les parece anticuada a

Harman, Keith Lehrer y a los pensadores anti-fundamentistas en general. Para ellos, la única clase de principio epistemológico general que podemos esperar es algo como lo siguiente:

> *S* está completamente justificado en creer que *p* si y solo si, dentro del sistema doxástico corregido de *S*, se cree que *p* tiene más probabilidades de ser verdadero que la negación de *p* o que cualquier otro enunciado que compita con *p*.[7]

En un principio así, todo gira en torno a qué es un «sistema doxástico corregido». Lehrer define esa noción del siguiente modo:

> El sistema *doxástico* de una persona *S* es un conjunto de enunciados de la forma *S* cree que *p*, *S* cree que *q*, etc., que describe lo que cree *S*. El sistema doxástico *corregido* de *S* es aquel subconjunto del sistema doxástico que resulta de borrar todos los enunciados que describen a *S* creyendo algo que dejaría de creer como buscador imparcial y desinteresado de la verdad […] como un investigador *veraz*.[8]

Lehrer prosigue explicando que «la forma en que un hombre averigua qué debería creer como investigador veraz es esmerándose en ese ideal y viendo qué es lo que cree».[9] Admite que «sigue siendo una posibilidad lógica que las creencias sobre el mundo que tiene una persona dentro de su sistema doxástico corregido sean enteramente erróneas»,[10] la tradicional objeción a las epistemologías coherentistas. Pero se contenta con dejarla sin respuesta, a diferencia de Descartes.

Comentando mi propia opinión sobre el papel de la epistemología en la filosofía moderna, Goldman ha insinuado que confundo la epistemología con una variedad específicamente cartesiana de ella, y que he tomado las críticas justificadas a la especie como críticas al género entero. Escribe:

> No hace falta suscribir el fundamentismo cartesiano para ser fundamentista. Incluso si no hay un «estar dada inmediatamente» de la mente, el conocimiento o la justificación pueden tener una estructura vertical. Por último, el fundamentismo no es la única teoría que hay para la justificación epistémica, el coherentismo es una opción viable.

> ¿Por qué no podría uno suscribir el coherentismo y seguir haciendo «epistemología»?[11]

La afirmación de Harman, que Goldman cita en «The Relation between Epistemology and Psychology», de que la lógica deductiva siempre puede llevarle a uno a cuestionar las premisas en vez de a adoptar las conclusiones, me parece que basta para mostrar que los casos donde «la justificación tiene una estructura vertical» son simplemente los casos en que nadie necesita asesoramiento epistemológico. Son esos en que todo el mundo ha aceptado ya jugar con ciertas reglas; a saber, reglas que prohíben cuestionar los datos o romper con el vocabulario o los principios tradicionales. En cuanto al coherentismo como una modalidad de epistemología, hay una diferencia clara entre la clase de cosas que esperan quienes buscan «reglas epistémicas correctas» y el tipo de estímulo moral que es todo lo que pueden ofrecer coherentistas como Lehrer y Habermas. Si lo que querías eran reglas para dirigir tu espíritu y lo que te dicen es que debes ser un buscador de la verdad imparcial y desinteresado que intente mantener sin distorsiones la situación de habla, probablemente te sentirás estafado. Así que yo diría que, una vez que suscribes el coherentismo, dudarás tanto de las reglas epistémicas como dudan de la noción de principio moral quienes han suscrito a Sartre.

Para precisar más esta cuestión, déjenme que pase ahora a la distinción de Goldman entre epistemología descriptiva, analítica y normativa. Coincido con él en que muchos manuales explican la epistemología de forma hipócrita. Uno no puede separar la epistemología de la descripción psicológica, al modo puritano habitual de Frege-Husserl, mientras sigue citando a Locke y Hume como distinguidos epistemólogos. Aplaudo que quiera incluir «la sociología de la ciencia y del conocimiento, la historia de la ciencia y la antropología cultural» como ramas de la «epistemología descriptiva social».[12] No obstante, yo añadiría a ellas la historia intelectual en general, incluyendo la obra de autores como Comte, Hegel, Marx, Nietzsche y Dewey. Son autores que describen los instrumentos conceptuales defectuosos que utilizaron nuestros

antepasados. De ahí sacan además moralejas normativas, avisándonos de los residuos suyos que aún quedan en nuestros vocabularios contemporáneos y de los que hacen presuposición nuestras instituciones culturales.

Dentro del campo de una epistemología descriptiva definida con esa amplitud, yo diría, sin embargo, que el lugar que puede ocupar la psicología —la epistemología descriptiva «*individual*», en palabras de Goldman— es más bien pequeño.[13] Si quieres ser un buscador de la verdad más desinteresado de lo que eres, si te preocupa caer preso de los prejuicios e ilusiones de tu tiempo como les ocurrió a los astrónomos aristotélicos de Padua en tiempos de Galileo, entonces a quienes querrás leer es a los epistemólogos *sociales* descriptivos. No parece probable que la psicología individual pueda hacer mucho. Por ofrecer una analogía a la que volveré más adelante, se puede entender que la historia intelectual —mi tipo favorito de epistemología social descriptiva—da pistas sobre cómo reprogramar a los organismos humanos para que sirvan mejor a los fines de la raza. La psicología puede verse entonces como el estudio del *wetware* —por desgracia inmutable— de esos organismos en la medida en que pone dificultades a tal reprogramación. Me parece que esta analogía ayuda a explicar por qué las esperanzas que el siglo XVII pensó que podrían cumplirse con la disciplina lockeana que ahora llamamos «psicología», en realidad las cumplieron disciplinas que no surgieron hasta los siglos XVIII y XIX. Fue en ellos cuando la conciencia histórica se convirtió en el poderoso motor de progreso intelectual que hoy reconocemos que es.

Volviendo tras esta prolepsis a la división tripartita de Goldman, dejo ya la epistemología descriptiva y paso a considerar la «epistemología analítica». Goldman describe ese campo como "interesado únicamente en el *concepto* de conocimiento y en el *concepto* de creencia justificada de la persona *común.*[14] Entiendo que está pensando, por ejemplo, en los intentos de proporcionar un cuarto criterio de conocimiento para resolver el problema de Gettier.

Simpatizo con la afirmación de William W. Rozeboom de que no hay razón para pensar que el uso de «saber» de la persona común

nos dé recursos para decidir si llamar o no «conocimiento» a casos de complicados rompecabezas gettierológicos, igual que su concepto de «moral» carece de recursos para resolver dilemas morales complejos. Así que dudo de que la epistemología analítica, en el sentido que define Goldman, tenga el menor futuro. La moda de hoy de dar condiciones necesarias y suficientes para la verdad de «S sabe que *p*» se me asemeja a la moda de siempre de darlas para la verdad de «lo correcto era que S hubiera hecho A». Dentro de los márgenes estilísticos de tales análisis se pueden traer a colación muchísimas cuestiones interesantes, pero pensar en un análisis que fuera a la vez informativo y a prueba de contraejemplos es una quimera. Como ha dicho Fodor, «la lección de la filosofía moderna es, con toda seguridad, que las ideas interesantes no se dejan analizar».[15]

Pasando ahora a la tercera división de Goldman —la epistemología normativa—, estoy de acuerdo con él en que aquí es donde está el lío. Yo argüiría que la esperanza de que cabe encontrar reglas epistémicas es la fuente original de la idea de que puede haber una disciplina llamada «teoría del conocimiento». Fue la esperanza de una epistemología normativa lo que hizo que aparecieran la descriptiva y la analítica. La esperanza era que podíamos volvernos buscadores más eficaces de verdades reprogramándonos, no simplemente con vocabularios nuevos en dominios particulares (como en el caso de Galileo), sino aislando lo que Goldman denomina «reglas independientes-del-dominio».[16] La cuestión que de verdad nos separa a él y a mí es la de si hay alguna probabilidad de extraer reglas independientes-del-dominio desde la psicología individual.

Mi pálpito de que no la hay procede de la reflexión de que las esperanzas de la epistemología solo se cumplirían si lográramos reglas que fueran, no solo independientes-del-*dominio*, sino independientes-de-la-*tarea*. Lo que entiendo por una regla independiente-de-la-tarea viene ejemplificado por la que está implícita en el ejemplo que Goldman cita de Lawrence Powers: cuando te pidan que adivines una palabra a la que le falta la primera

letra, utiliza imágenes visuales más bien que auditivas. Otro ejemplo sería: cuando inventes nuevos chistes «toc-toc»[17] para entretener a los niños, prueba con diferentes imágenes auditivas más que depender de las visuales. Supongo que Goldman admitiría que no es esta la clase de reglas que quiere la epistemología normativa. Pero se diría que es la única que la psicología parece poder darnos. La razón por la que lo creo quizá no pueda expresarse mejor que recuperando la analogía que mencioné antes: la de la psicología con las cuestiones de hardware, por un lado, y de la historia intelectual con las cuestiones de programación, por otro. Pensemos en un analista de sistemas a cargo de una sala llena de computadoras (todas del mismo modelo). Está intentando hacer que las computadoras sirvan mejor a los intereses de la empresa. Pasará la mayor parte del tiempo elucubrando nuevas formas de presentarles los datos, nuevos problemas para que los resuelvan, nuevos lenguajes con que programarlas, etc. Necesitará saber muy poco sobre su hardware, salvo para pequeñas tareas muy específicas que aparecen en el curso de la programación. Puede sacar partido de algún detalle ingenieril para ahorrar un poquito de tiempo de procesamiento, o puede que tenga que complicar un programa para sortear algún otro detalle ingenieril que alguien tuvo la pésima idea de incorporar en sus máquinas. Pero, cuando se hace preguntas amplias sobre qué operaciones es más probable que conduzcan al tipo de respuestas que quiere obtener, tirará de su experiencia con todos los programas y formatos descartados anteriormente, de las historias de éxitos y fracasos que cuentan sus colegas analistas. Ese acervo será su equivalente de la historia intelectual. Probablemente no pensará que más detalles sobre el modo de funcionamiento del modelo al que está atado —su equivalente de la psicología individual — vayan a ayudarle.

No obstante, alguien podría sugerir que se le puede dar la vuelta a la analogía para que opere a favor de Goldman. Pues, mientras que es de suponer que el analista de sistemas tiene al lado un manual con los detalles ingenieriles de sus máquinas, nosotros todavía estamos tratando de averiguar los detalles de la máquina

que es nuestro cuerpo. Se puede interpretar que la psicología cognitiva descubre el programa del organismo humano, o más bien, los programas de los diversos mecanismos relativamente independientes que lo componen. Por tanto, no hay razón a priori para pensar que, una vez que los manuales de nuestro organismo estén escritos, no vayamos a sacar de ellos reglas epistémicas independientes del dominio y de la tarea, instrucciones para reprogramarnos. Con todo, hay una razón a posteriori para dudarlo: que la mayoría de los resultados interesantes de experimentos psicológicos que citan los epistemólogos y los filósofos de la mente están sencillamente demasiado por debajo del nivel de abstracción en el que se mueve el epistemólogo, el nivel en el que podrían formularse reglas independientes del dominio y de la tarea. Esto es algo que dice Daniel Dennett en su reseña de *El lenguaje del pensamiento*, de Fodor. Dennett apunta que, aunque de un programa de ajedrez uno pueda comentar correcta y provechosamente que «piensa que debería sacar pronto la reina»,

> en ninguna parte de los muchos niveles de representación explícita que tiene el programa hay explícitamente simbolizado algo más o menos sinónimo de «debería sacar pronto la reina». El nivel de análisis al que pertenece […] el comentario describe rasgos del programa que son, dicho de manera enteramente ingenua, propiedades *emergentes* de los procesos computacionales que tienen «realidad ingenieril».[18]

Dennett recapitula diciendo que «no veo razón para creer que la relación entre el modo de hablar en términos de creencias y el modo de hablar en términos de procesos psicológicos no haya dejado ya para siempre de ser directa».[19] Goldman, me parece a mí, tendría que satisfacer las dudas de Dennett dándonos un ejemplo de algún fragmento de investigación psicológica que prometa decirnos algo sobre estrategias intelectuales generales.

En este punto, creo que una comparación histórica puede resultar instructiva. Que Hume pusiera al descubierto el hiato que se abría entre el modo de hablar lockeano de asociaciones de ideas y el modo de hablar científico de sustancias y leyes causales fue lo que

llevó a Kant a concluir que la psicología —lo que él llamaba «la fisiología de Locke del entendimiento humano»— no iba a servirle de mucho a la epistemología normativa. Kant —el epistemólogo normativo *par excellence*— quiso separar los programas de investigación prometedores de los no prometedores dentro de la cultura de sus días. Sospechaba que la física newtoniana estaba perfectamente, que la biología vitalista era dudosa y que la metafísica leibniziana no tenía arreglo. Para argumentar estas sospechas le pareció que era necesario transformar la epistemología, pasando de ser el intento lockeano de describir transiciones de estado mental a estado mental a ser algo completamente distinto. Para él, devino en una serie de experimentos mentales concebidos para mostrar qué ocurriría si uno intentara sobrepasar lo que él consideraba que eran los límites del conocimiento humano. Dichos experimentos mentales le dieron razones para decir, por ejemplo, que la metafísica leibniziana jamás iba a alcanzar la senda segura de una ciencia.

La transformación de la epistemología por parte de Kant alumbró rápidamente, en Herder y en Hegel, los comienzos de lo que Goldman llama «epistemología descriptiva social», la tentativa de concebir formas nuevas de vida intelectual que quizá pudieran ampliar esos límites. Como dije antes, la clase de historia intelectual y de antropología cultural que con ello fundaban estos sucesores de Kant desplazó en buena medida a la psicología (no digamos ya a la epistemología) como fuentes de asesoramiento para los intelectuales que quieren saber cómo dirigir su espíritu. Para volver a poner a la psicología descriptiva *individual* lockeana —el estudio de las transiciones entre estados mentales— en el lugar que tenía antes de Kant, y así darle a Goldman lo que necesita, los psicólogos tendrán que salir con algo que parezca un verdadero avance respecto de Locke. Prácticamente las únicas candidatas a eso que le vienen a uno a la cabeza son las teorías freudianas de la motivación inconsciente. Pero tales teorías parecen, si acaso, agua para el molino de historiadores intelectuales como Foucault antes que surtidores de ejemplos para Goldman. Es cierto que los doctorados en psicología tienen a la venta un repertorio de trucos con

descubrimientos freudianos que pretendidamente valen para la gestión de personal, la asesoría matrimonial y similares. Pero, una vez más, son trucos que parecen demasiado dependientes-de-la-tarea para desempeñar el papel de las «reglas epistémicas» de Goldman.

Puedo resumir mi reacción al artículo de Goldman del siguiente modo. Estoy de acuerdo en que, dados los principios meta-epistemológicos que él postula, la psicología es relevante para la epistemología. Además, considero que demuestra muy bien el fracaso de diversos intentos de casar a la epistemología con la lógica y excluir a la psicología. Muchas de las cosas concretas que dice sobre qué estaría bien que tuviéramos en el ámbito de la epistemología me parecen acertadas y muy claramente razonadas. Pero sigo sin poder convencerme de que haya reglas epistémicas del tipo que sus principios meta-epistemológicos exigen. Por eso, para mí su demostración de que la psicología es relevante para la epistemología se parece a una que demostrara que la física de partículas es relevante para la alquimia. No veo razón para creer que necesitemos revisar la decisión que tomaron los intelectuales a comienzos del siglo XIX de buscar en la historia y no en la psicología, en la epistemología descriptiva social y no en la individual, las pistas sobre cómo pensar.

* * *

Pasando ahora de Goldman a Fodor, mis dudas sobre su artículo[20] son del mismo nivel de generalidad que las que me suscita el de Goldman. Afectan al proyecto como un todo más que a los detalles de su ejecución. Pongo en duda la idea de que podamos explicar «las propiedades intencionales de los estados mentales como heredadas de las propiedades semánticas de las representaciones mentales que [...] intervienen en su simbolización».[21] Más en concreto, no tengo claro qué significa ahí «heredadas de». Las propiedades intencionales de los estados mentales, dice Fodor, incluyen su ser acerca de algo y el tener un valor de verdad. Para tener la impresión de que Fodor ha resuelto un problema real, uno

necesitaría imaginar a alguien a quien le produjera perplejidad que los estados mentales puedan poseer esas propiedades. Además tendría que ser alguien que sintiera que su perplejidad se desvanece cuando le dicen que podemos seguir el procedimiento que esboza Fodor hacia el final de su artículo[22] —encontrar condiciones básicas de una representación mental que sean a la vez el contenido de la creencia del sujeto y la condición de verdad de la representación— y dar así semanticidad a las representaciones mentales. O sea, uno tiene que imaginar que el auditorio al que se dirige Fodor son filósofos que creen las siguientes cosas:

(a) No basta con entender la intencionalidad de los estados mentales para entender, a la manera de Ryle, cómo verificamos las adscripciones de tales estados.
(b) Basta entender la semanticidad de las representaciones mentales para entender, a la manera de Fodor, cómo verificamos las adscripciones de condiciones de verdad a esas representaciones.
(c) Entender este segundo proceso de verificación legitima de algún modo el primero porque ahora podemos ver la intencionalidad de los estados como heredada de la intensionalidad de las representaciones.

Yo no conozco a ningún filósofo así. Me parece que a estas alturas a nadie le impresiona ya la tesis de Brentano: que no hay condiciones de comportamiento necesarias y suficientes para la verdad de la adscripción de estados intencionales. Es así porque todos hemos asimilado lo que Fodor llama «la lección de la filosofía moderna», la de que no hay condiciones necesarias y suficientes de nada que sean interesantes. Admito que existen filósofos que, aunque ya no suspiran por bicondicionales, aún tienen la impresión de que la intencionalidad es algo profundo y misterioso. Pero no creo que esos filósofos sientan que Fodor ha resuelto el misterio. Searle y Dreyfus, por ejemplo, creen que hay algo llamado «intencionalidad intrínseca» que solo entenderemos si logramos entender mucho más sobre el cerebro de lo que entendemos ahora.

Dreyfus ve la búsqueda de la intencionalidad intrínseca como la continuación de la insistencia de los fenomenólogos en la importancia e «irreductibilidad» de la intencionalidad. Pero ni Dreyfus ni Searle pensarían que el secreto de la intencionalidad intrínseca habría sido desvelado al reparar en que podemos adscribir condiciones de verdad a sucesos neurales de la forma que propone Fodor. Replicarían que podemos adscribir esas condiciones de verdad a los estados internos de las computadoras —o, para el caso, de los termostatos— igual de fácilmente, y que el que podamos hacerlo no arroja luz sobre «la naturaleza de la intencionalidad». Porque, para ellos, que las computadoras y los termostatos no tienen estados intencionales es un principio de la ciencia cognitiva que no se cuestiona. Explicar que, asignando condiciones de verdad a sucesos internos de esos aparatos, uno puede tratarlos como si contuvieran representaciones internas del entorno, a Dreyfus y Searle les dejaría fríos.

Si se les dice que los termostatos pueden «heredar» intencionalidad de la semanticidad de algunos de sus estados internos, tendrán prejuicios exactamente igual de fuertes para aceptar que esos estados internos son verdaderamente semánticos, que los que tienen para aceptar que el termostato como un todo verdaderamente tiene creencias.

Permítanme que esboce lo que para mí es el cuadro del escenario actual en filosofía de la mente que constituye el telón de fondo dialéctico de los esfuerzos de Fodor. En el ala izquierda, hay gente como Dennett y yo que adscribimos alegremente ser-acerca-de y tener valor de verdad a cualquier cosa que resulte práctico describir en términos intencionales. Tal descripción es ciertamente práctica para personas y computadoras, práctica pero opcional para termostatos y gramolas, y con toda probabilidad muy útil para estados de los cerebros. Como dice Fodor, la única idea que hemos encontrado para hacer psicología no-conductista es una teoría de las representaciones internas, y como la psicología conductista nunca dio mucho de sí, no parece que haya nada de malo en darle cancha a la psicología cognitiva. Entonces, diría Dennett, sigamos adelante y adscribamos «contenido» a estados en el nivel «subpersonal». Pero

no nos preocupemos de si tenemos alguna justificación filosófica para hacerlo. Newton postuló la acción a distancia a pesar de que le gritaban que estaba resucitando las cualidades ocultas de los escolásticos. Compensó. Quizá compense adscribir condiciones de verdad a los sucesos cerebrales o quizá no compense, pero, en cualquier caso, podemos ignorar tranquilamente los gritos de los filósofos.

En el ala derecha, está la gente entregada a esa clase de gritos, Dreyfus y Searle. Nos ven a Dennett y a mí como verificacionistas que confundimos la sombra de la intencionalidad —el comportamiento que Dennett aduciría como evidencia de la intencionalidad del termostato— con su sustancia. Ven una diferencia real entre la «intencionalidad intrínseca» que tienen las personas y los perros y la «pseudo-intencionalidad» que tienen las máquinas. Presumiblemente verían una diferencia paralela entre la semanticidad intrínseca que posiblemente tengan los sucesos de un cerebro y la pseudo-semanticidad de los sucesos de un transistor. Así que dirían que Fodor ha abonado el terreno para el verificacionismo al desmitificar la semanticidad de los estados cerebrales. Desde el punto de vista de esta ala derecha, la pregunta que hay que contestar no es «¿cómo empezar a adscribir condiciones de verdad a estados cerebrales?», sino «¿qué tienen los cerebros que no tienen los transistores y que hace que no debamos adscribir condiciones de verdad a los estados de las máquinas?» Fodor elude esta pregunta, dirían ellos, dándonos procedimientos para adscribir semanticidad que funcionarían con cualquier cosa con un grado razonable de complejidad.

Fodor se sitúa entremedias, con simpatías hacia el ala derecha y hacia la izquierda. Cree que Dennett es demasiado frívolo e insuficientemente «realista» a la hora de asignar intencionalidad. Coincide con Searle y Dreyfus en que

> ser una instancia del mismo programa que el del cerebro no es, en y por sí mismo, una condición suficiente para tener aquellas actitudes proposicionales características del organismo que tiene un cerebro.[23]

Fodor piensa que debe cumplirse además la condición de que haya «el tipo correcto de enlazamiento causal», el mismo «que rige entre nuestro cerebro y nuestros mecanismos transductores, por un lado, y entre nuestro cerebro y los objetos distales, por otro».[24] Se supone que también piensa que el tipo correcto de enlazamiento causal es lo que velaría por la «herencia» de la intensionalidad desde los sucesos simbólicos en el cerebro al estado del organismo. Así que tal vez Fodor responda a Searle y Dreyfus que de lo que deberían quejarse no es de cómo analiza él lo que se requiere para que un suceso interno tenga una condición de verdad, sino de que no especifica qué enlazamientos causales tienen los cerebros y de los cuales carecen las computadoras. Fodor da a entender que sencillamente no sabemos si hay algo que no sean cerebros que presente el tipo correcto de relación causal. Así que podría decir que desmitificar lo que Searle y Dreyfus llaman «intencionalidad intrínseca» requiere dos pasos, de los que él solo ha dado uno. El primer paso es mostrar cómo pueden tener condiciones de verdad los sucesos internos; el segundo es mostrar qué relaciones causales constituyen la relación de «herencia» que enlaza la semanticidad de las representaciones con la intencionalidad de los estados. Acto seguido podría admitir encantado que este segundo paso aún está por dar.

Sin embargo, no sé muy bien cómo proseguiría el debate entre la posición centrista de Fodor y la derechista de Searle, pues el uso que hace Fodor de «herencia» me resulta tan vago como el que Searle hace de «intrínseca». Fodor afirma que la estrategia de «reducir el problema "¿qué hace intencionales a los estados mentales?" al de "¿qué confiere propiedades semánticas a (fija la interpretación de) un símbolo?"»[25] se remonta a Locke. Dice que Locke también sostuvo que «las propiedades intencionales de los estados mentales se heredan de las propiedades semánticas (referenciales) de las representaciones mentales».[26] Puede que sí, aunque no estoy seguro de cómo reformular el punto en la jerga de Locke. Para él, después de todo, el rasgo definitorio de una «idea» era que esté «ante la mente», no que tenga semanticidad.

Posiblemente Locke pensaba que las ideas tenían otros rasgos que de alguna forma las comunicaban con las mentes que las tenían, pero no recuerdo nada en Locke que explique en qué consiste esa comunicación. De modo que sospecho que, si Locke realmente sostuvo esa posición, su versión de ella necesita tanta explicación como la de Fodor.

Quizá estoy apretando demasiado a Fodor por el uso que hace de «heredar», pero insisto en que nos debe una explicación más completa de la naturaleza del problema «qué hace intencionales a los estados mentales» antes de que podamos evaluar su propuesta de reducirlo al de la semanticidad de las representaciones internas. Para izquierdistas como Dennett y yo, no existe ningún problema con esa forma. Sabemos todo ya sobre lo que hace intencionales a los estados mentales: de ellos se dice que son sobre cosas y que tienen valores de verdad. A nuestra manera superficial y pragmática, la pregunta «¿cómo puede algo tener tan extrañas propiedades?» no nos preocupa porque tener un valor de verdad no parece más extraño que ser rojo o ser bueno, o que tener peso atómico. A los izquierdistas, la pregunta «¿los termostatos tienen realmente creencias y los estados cerebrales condiciones de verdad, o es simplemente que decirlo resulta cómodo?» nos parece tan chocantemente tonta como la pregunta «¿las cosas son realmente rojas, o es simplemente que resulta cómodo usar el predicado "rojo" para describirlas y explicarlas?»

Sospecho que el problema subyacente entre Dennett y yo, por un lado, y Fodor, por el otro, es justamente la cuestión de si hay alguna diferencia interesante entre verdad y conveniencia. En una crítica a «Intentional Systems»,[27] de Dennett, Fodor le cita cuando dice que «la decisión de adoptar la estrategia (de instalarse en la actitud intencional) es pragmática y no es correcta o incorrecta intrínsecamente», y lo glosa así: «el análisis de Dennett […] explica la utilidad […] del lenguaje intencional sin dar por supuesto que haya hechos a los que se correspondan las adscripciones intencionales».[28]

Mi problema aquí es que no veo qué más podría hacer uno en términos de «dar por supuesto que hay hechos a los que se corresponden las actitudes intencionales» sino explayarse sobre la utilidad de las adscripciones. Además, dice Fodor que

> si empezamos por pensar en nuestras teorías de actitud intencional (por ejemplo, en nuestra lógica) como simples dispositivos computacionales de los que se extraen de algún modo predicciones sobre fenómenos mentales, entonces es seguro que se suscita la pregunta «¿qué autoriza la transición desde relaciones entre fórmulas dentro de la teoría a relaciones causales en la mente?»[29]

Yo no veo que esa pregunta se suscite. Pero, incluso si fuera así, no veo que la respuesta pudiera ser la que Fodor recomienda, a saber, que «es el supuesto de que la teoría [de actitud intencional] es *verdadera* ».[30] Eso no autoriza la transición, simplemente afirma la intención de hacerla.

Probablemente Fodor replicaría que mi visión de la verdad es incorrecta. Cuando él dice que «la teoría es verdadera», sigue explicando, se refiere a que, «en algo parecido en espíritu al sentido técnico, los procesos mentales del agente constituyen un modelo de la teoría».[31] Fodor cree que, al decir algo como eso, has dado una razón para usar una teoría que añade algo a tu esperanza o tu creencia de que la teoría funciona, de que realmente te permite predecir y controlar. Yo no lo creo. Además, Fodor piensa que la carga de la prueba para dar «un argumento de que la interpretación Realista no puede sostenerse»[32] recae en personas como Dennett y yo. Aquí estamos en un impasse dialéctico. Para mí, si es que no para Dennett (quien a veces, qué pena, expresa anhelos Realistas que traicionan su enfoque notablemente pragmático), corresponde a Fodor y a otros Realistas explicar qué clase de razón adicional has dado para usar una teoría cuando dices «y además entiendo que se corresponde con la realidad». A mí eso me parece como apuntalar tu creencia de que el opio te hará dormir recordándote a ti mismo que tiene poderes dormitivos.

Déjenme que intente conectar ahora este impasse con la cuestión de si existe algún problema sobre «qué hace mentales a los estados intencionales». Yo no veo un problema ahí porque traduzco la pregunta a la pregunta ryleana «¿qué procedimientos de verificación tenemos para adscribir creencias y deseos a los organismos?» Fodor no hará esa traducción porque cree que las preguntas sobre la naturaleza de algo —por ejemplo, de la intencionalidad— persisten después de que hayamos dicho todo lo que tenemos que decir sobre procedimientos de verificación. Esa es la opinión que uno esperaría de un Realista, y la mía es la que se esperaría de un pragmatista. Como este segundo y más amplio debate es demasiado amplio para tenerlo aquí, lo dejo en este punto.

Terminaré, no obstante, señalando un modo más general en que el debate realismo vs. pragmatismo es relevante para nuestras discusiones. Considérese la pregunta: ¿qué tiene que ver con la epistemología el intento de Fodor de responder a la pregunta «qué hace intencionales a los estados mentales»? O, en términos más generales, ¿qué tiene que ver la pregunta «cuál es la naturaleza de la mente» con la de «cuál es la naturaleza del conocimiento»? La respuesta Pragmatista a esta última es: seguramente no mucho. O sea, el Pragmatista sospecha que todo lo que hay que decir sobre el conocimiento queda dicho cuando estudiamos los procedimientos de verificación, los juegos de lenguaje que practican los investigadores. Como dije antes, esa es la tarea de lo que Goldman llama «epistemología descriptiva social»: la historia intelectual y similares. Así que al Pragmatista le da igual —por lo que hace a la epistemología— si la mente es una sustancia espiritual cartesiana, o un haz de estados cerebrales con propiedades semánticas, o solo un manojo de poleas y palancas sin propiedades semánticas. En cambio sospecho que, para un Realista como Fodor, decir «el conocimiento solo es posible porque tenemos representaciones internas de la realidad» resulta epistemológicamente iluminador. Para el Pragmatista, sin embargo, no lo es más que decir «el conocimiento solo es posible porque tenemos un sistema nervioso». El Pragmatista ve el conocimiento como algo que estudias en el nivel más bien elevado de la programación, no en el del *wetware*, ni siquiera el del

wetware considerado como dispositivo computacional. Por las razones que sugerí antes, los detalles del *wetware* solo parecen trozos últimos en que el conocimiento probablemente puede dividirse. Así que el Pragmatista le desea al psicólogo cognitivo que le vaya bien con la Teoría Representacionista de la Mente, como se lo desea a Newton con la acción a distancia. Pero no espera más iluminaciones epistemológicas de uno que iluminaciones metafísicas del otro.

---

[1] René Descartes, *Discourse on Method; and Meditations on First Philosophy*, 4.ª ed., trad. de Donald A. Cress. Indianápolis, Indiana, Hackett, 1998, II. —Eds. [N. del T.: seguimos la traducción de Guillermo Quintás, *Discurso del método. La dióptrica. Los meteoros. La geometría*. Barcelona, Círculo de Lectores, 1996.]

[2] Véase Alvin I. Goldman, «Epistemics: The Regulative Theory of Cognition», *The Journal of Philosophy* 75, n.º 10 (1978), pp. 509-523. —Eds.

[3] Marshall Swain, «Epistemics and Epistemology», *The Journal of Philosophy* 75, n.º 10 (1978), p. 524.

[4] Véase Richard Rorty, *Philosophy and the Mirror of Nature*. Princeton, Nueva Jersey: Princeton University Press, 1979, capítulo 3. —Eds. [Hay trad. esp.: *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, 1983.]

[5] Alvin I. Goldman, «The Relation between Epistemology and Psychology», *Synthese* 64, n.º 1 (1985), p. 36.

[6] Gilbert Harman, *Thought*. Princeton, Nueva Jersey: Princeton University Press, 1973, p. 157.

[7] Keith Lehrer, *Knowledge*. Oxford: Clarendon, 1974, p. 198.

[8] *Ibid*., pp. 189-190.

[9] *Ibid*., p. 190.

[10] *Ibid*., p. 213.

[11] Alvin I. Goldman, «Review of *Philosophy and the Mirror of Nature*, by Richard Rorty», *The Philosophical Review* 90, n.º 3 (1981), p. 425.

[12] Goldman, «The Relation between Epistemology and Psychology», p. 31.

[13] *Ibid*.

[14] *Ibid*., p. 33.

[15] Jerry A. Fodor, «Psychosemantics; or, Where Do Truth Conditions Come From?», en *Mind and Cognition*, ed. por William Lycan. Oxford: Basil Blackwell, 1990, p. 333. —Eds.

[16] Goldman cree, a despecho de Kuhn, que las diferentes disciplinas *sí* se dotan de cánones metodológicos, de reglas independientes-del-dominio. Piensa que la epistemología debe suministrar cánones aún más generales. Yo me alineo con Kuhn en pensar que las diversas disciplinas proporcionan meramente metodologías de saber-hacer, no metodologías independientes-del-dominio. Pero es esta una cuestión colateral en la que no hace falta entrar a los presentes propósitos.

[17] N. del T.: «*knock-knock jokes*». Los típicos chistes infantiles con un diálogo que empieza «toc-toc / ¿quién es?» y se resuelven con algún juego fonético. Por ejemplo: toc-toc; ¿quién es?; Lola; ¿qué Lola?; Lola-drones.

[18] D. C. Dennett, «Critical Notice», *Mind* 86, n.º 342 (1977), p. 279.

[19] *Ibid*.

[20] Fodor, «Psychosemantics; or, Where Do Truth Conditions Come From?» —Eds.

[21] *Ibid*., p. 313.

[22] Cf. *ibid*., p. 332.

[23] J. A. Fodor, «Searle on What Only Brains Can Do», *Behavioral and Brain Sciences* 3, n.º 3 (1980), p. 431.

[24] *Ibid*.

[25] *Ibid*

[26] *Ibid*.

[27] Daniel C. Dennett, «Intentional Systems», *Journal of Philosophy* 68, n.º 4 (1971), pp. 87-106. —Eds.

[28] Jerry A. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, Massachusetts: MIT Press, 1981, p. 105.

[29] *Ibid*., p. 113.

[30] *Ibid*., p. 114.

[31] *Ibid*.

[32] *Ibid*.

CAPÍTULO 11
LA OBJETIVIDAD DE LOS VALORES

I

Mucha gente cree que los filósofos deberían poder decirte qué es real y qué no, como si hubiera algún modo filosófico especial de enmendarle la plana a la ciencia y al sentido común y llegar a la verdadera naturaleza de las cosas. Muchos creen también que es parte del trabajo del filósofo decirnos si son reales los valores; si palabras como «correcto» y «bueno» se refieren a propiedades «objetivas» de las situaciones o las acciones, o más bien a algo que es meramente «subjetivo» o «relativo», como reacciones emocionales o costumbres sociales. A mí esta idea de lo que deben hacer los filósofos me parece totalmente equivocada. No creo que dispongan de métodos especiales para decidir que la ciencia tiene razón y la religión no, o viceversa, ni para dirimir conflictos entre la ciencia y el sentido común. Lo único que pueden hacer es tratar de encontrar un vocabulario en el que sea posible ver cómo encaja todo entre sí: la ciencia, la religión, el sentido común, la moralidad. Dicho de otro modo, todo lo que pueden hacer es intentar limar conflictos entre esas áreas de la cultura de forma que podamos conservar tanta religión como sea posible en vista de la ciencia, tanta ciencia como sea posible en vista del sentido común, y así sucesivamente. Por decirlo aún de otro modo más, los filósofos no pueden salir con nuevos descubrimientos sorprendentes de que Dios no existe, o de que la moralidad es una ilusión, o de que la ciencia moderna es solo una distorsión «conceptual» de la realidad. Dicen cosas de ese tipo, eso seguro, pero debemos verlas como eslóganes, como dramatizaciones de su nueva forma de mirar a las cosas. No han hecho un descubrimiento, en el sentido en que un astrónomo puede descubrir un nuevo planeta o un químico sintetizar un nuevo compuesto. Ni han demostrado nada, en el sentido en que un

matemático puede demostrar un teorema o un abogado la validez de un contrato. Sus anuncios son en realidad solo anuncios de un nuevo vocabulario, una nueva perspectiva, una propuesta sobre cómo reformular lo que hemos estado diciendo en unos términos que nos darán una visión más coherente, menos expuesta a objeciones, que confiera más sentido a las diversas cosas de las que hablamos.

Si esta visión de la filosofía es correcta, entonces la idea de que los filósofos deben decirnos si los valores son o no reales va descaminada. «¿Son los valores objetivos o subjetivos?» es una mala pregunta. Hay montones de cosas que decir sobre las relaciones entre los valores morales y las emociones, los valores morales y Dios, los valores morales y las costumbres sociales, etcétera, pero es una tontería pensar que los filósofos puedan descubrir que los valores morales son reacciones emocionales. Tampoco tiene sentido imaginarse leyendo en el periódico que los filósofos han descubierto, para sorpresa de todos, que eran tan reales y objetivos como los planetas o los compuestos químicos. Ni van a descubrir tampoco valores morales nuevos, inesperados, en los que nadie había reparado hasta ahora pero que cambian por completo la política social o el modo en que uno debe vivir su vida.

Por formular mi tesis de otra manera, hablar de valores morales es emplear una jerga errónea para hablar de lo que la gente debe hacer: cómo deben vivir, cómo deben tratar a otras personas, cómo podrían sopesar obligaciones en conflicto, cómo pueden organizarse en sociedades. Describir las decisiones sobre tales materias como si se hicieran mirando a algo llamado «el reino de los valores», o estudiando algo llamado «teoría de los valores», o estableciendo rankings relativos entre valores diversos, es una metáfora desafortunada. Suena como si las vidas, o las acciones, o las situaciones, vinieran con una pequeña etiqueta que te dijera cuánto de cada valor ejemplifica cada una (como otras etiquetas detallan la composición de una fibra o el contenido nutricional). Suena como si el único problema fuera clasificar en su orden esos valores. Pero si concluyo que el contenido de Goethe era alto en autorrealización y bajo en altruismo, o que el de Lenin era alto en audacia, altruismo y

crueldad, y bajo en caridad y piedad, no estoy poniendo más fácil decidir si se adopta a esos hombres como modelo. Si digo que matar a un enfermo terminal agonizante sería alto en eliminación de sufrimiento, pero bajo en respeto por la vida humana, no estoy haciendo progresos para decidir si matar o no. Si digo que la sociedad china es alta en disciplina y en atención a lo social, y que la norteamericana lo es en libertad y tiempo libre, no estoy avanzando nada en la decisión de cómo debería organizarse la sociedad. Deliberar sobre asuntos morales no es cuestión de aislar propiedades morales especiales y luego clasificarlas, o, por lo menos, no debería serlo. La jerga del filósofo sobre «valores» nos hace más difícil recordar cómo es la deliberación moral y política en la vida real.

Es solo si tenemos esa imagen confundente de los valores como etiquetitas pegadas a las cosas, cuando se nos ocurre preguntar si las etiquetas están realmente ahí o solo en nuestra mente, o si pertenecen real y verdaderamente a las cosas o más bien es la cultura o la sociedad a la que pertenecen quien se las pega. Si, en lugar de ello, pensamos en el problema de cómo lograr una sociedad decente, o cómo vivir de manera que uno pueda respetarse a sí mismo, o cómo ayudar a alguien atrapado en un dilema moral, todo eso de la «objetividad» parece que deja de venir a cuento. Digan lo que digan los filósofos sobre el estatus metafísico de los valores o sobre el significado de las oraciones que se usan para hacer juicios morales, los problemas morales siguen siendo tan reales y apremiantes como siempre.

A mi modo de ver, la mala pregunta sobre la objetividad de los valores se plantea por una doble razón. En primer lugar, heredamos de los siglos XVII y XVIII la idea de una oposición entre «ciencia», esclarecida y severa, y «prejuicio» o «superstición», aturullado y blando[1]. Por consiguiente, creemos que es necesario ser «científicos» sobre la moral, y que eso significa tener claro si los valores morales están «ahí afuera» o existen únicamente «en nosotros».

En segundo lugar, desde que Trasímaco retó a Sócrates a explicar por qué la justicia compensa —por qué a uno le beneficia que le importe lo que les pasa a otros—, ha parecido que solo podíamos contestarle si algo como la Forma del Bien de Platón era real. Los filósofos dan a entender a veces que, salvo que la naturaleza del mundo sea al menos tan diferente del sentido común como Platón decía, no hay respuesta a Trasímaco o a Hitler, a la persona a la que sencillamente le da igual lo que les ocurra a los demás.

Me parece que la idea del siglo XVII de que aquello que la ciencia no conoce no es real es un error. También me lo parece la de Platón de que, a menos que los filósofos puedan descubrir un tipo nuevo y especial de objeto —por ejemplo, un reino de las Formas, un reino de valores—, no podemos responder a Trasímaco. En el resto de este artículo trataré de explicar por qué creo que son errores y, por ende, por qué considero que la pregunta sobre «la objetividad de los valores» es una mala pregunta. En la siguiente sección, sostendré que es absurdo distinguir entre cualidades de los objetos que están «en ellos» por oposición a las que están «en nosotros». En la sección 3, criticaré algunos de los argumentos de un reciente defensor de la idea de que no hay valores objetivos, el difunto J. L. Mackie, de Oxford. En la sección 4, me ocuparé de una variante de la noción de «valor moral objetivo», la de «principio moral». Ahí sostendré que, aunque no hay forma de «responder» a Trasímaco o a Hitler por deducción a partir de principios, eso en nada demuestra que la moralidad sea «subjetiva», o «personal», o «relativa». Por último, en la sección 5 argüiré que la idea de «filosofía moral» está obsoleta; que la novela, por un lado, y las ciencias sociales, por otro, la han remplazado en nuestro siglo como alternativa secular a la religión.

II
La distinción subjetivo-objetivo

La idea de que, aunque las mesas son puntiagudas y duras realmente, realmente no tienen color, mucho menos belleza o fealdad, se remonta como mínimo a Galileo. Pero si preguntamos

qué significa decir que una mesa es objetivamente dura y solo subjetivamente marrón o fea, no es fácil responder. Una forma de explicar la distinción es diciendo que permanece dura en todas las condiciones, pero es marrón solo a cierta luz y para ciertas personas (por ejemplo, las que no tienen ictericia). Pero eso no funciona. ¿Qué significa decir que permanece dura en todas las condiciones? Supongamos que uno vierte sobre ella algún mejunje que la reblandece, ¿por qué eso no cuenta como un cambio de condiciones? Imaginemos, alternativamente, que aparece alguien pidiendo algo duro y uno le sugiere que pruebe con la mesa. Supongamos que explica que, como lo que quiere es cortar cristal, necesita algo verdaderamente duro, como acero al carbono o un diamante. Bueno, podría decirse, al menos la mesa tiene una determinada dureza en un determinado momento, pero mucha resistencia a la presión en este preciso instante (antes de verter en ella el reblandecedor). ¿Quizá eso demuestra que su dureza es una propiedad objetiva, intrínseca, real, de la mesa? Bueno, supongamos que uno restringiera las cosas en el caso del color y dijera que en este preciso momento, a esta luz concreta, la mesa es marrón intrínsecamente, objetivamente. Se podría replicar que seguiría sin verse marrón para las personas cuyos ojos están en un cierto estado. Pero eso parece como decir que la mesa no sería dura si se calibrara con un medidor de presión que estuviera estropeado. Por supuesto que, en condiciones anormales, las cosas no parecerán tener sus propiedades normales. Pero ¿qué bases tenemos para dividir las propiedades normales en dos clases, la de las intrínsecas y la de las extrínsecas, unas «primarias» y las otras «secundarias»?

Sugiero que la razón por la que tendemos a hacer eso es que confundimos dos preguntas: (i) ¿cómo de debatible es la atribución de la cualidad a la mesa? y (ii) ¿la cualidad está en la mesa o en nuestra mente? La primera es totalmente razonable, y cabría decir que la dureza de la mesa, su color marrón y su fealdad se pueden disponer en ese orden dentro de una escala de debatibilidad. Armar una discusión sobre cuán dura es una cosa es muy difícil, es un poquito menos difícil tenerla sobre su color, y sobre cómo de fea es resulta considerablemente más fácil. Pero, cuando hacemos

equivaler debatibilidad creciente con creciente probabilidad de ser «meramente subjetiva», ¿qué estamos añadiendo? La idea de estar «en la mente» es obviamente metafórica, nadie piensa que tengamos una cajita inmaterial dentro del cráneo llena de objetos como la marronez o la fealdad. Pero si intentamos canjear la metáfora, no parece que recibamos nada. Decir que las opiniones sobre lo que es bello, o correcto, o bueno, varían en función de la educación de la gente, que se trata de asuntos particularmente debatibles, eso no lo discute nadie. Explicar ese hecho diciendo que son asuntos intrínsecamente subjetivos parece inútil. Pensemos en las distintas visiones que tienen las culturas sobre cómo es el universo: si está lleno de demonios, si está sujeto a leyes mecánicas, si su naturaleza está esencialmente vinculada al destino humano, si es finito o infinito, si está bajo el control de un creador benevolente, etcétera. Es obvio que la opinión que tenga uno sobre estos asuntos variará según su cultura, educación y demás. ¿Debemos explicar esa variación diciendo que no hay una verdad objetiva sobre la estructura del universo? Si es así, quizá sea mejor abandonar por las buenas la noción de objetivo vs. subjetivo porque una de sus mitades no se va a aplicar nunca a nada. Si no es así, ¿qué criterios tenemos para decir que las teorías científicas son sobre cuestiones de hecho objetivas mientras que las teorías morales no?

Creo que no hay una buena respuesta para esta última pregunta ni, por consiguiente, un uso para el que la distinción subjetivo-objetivo sea buena tampoco. Haríamos bien en remplazarla por una más clara entre aquello de lo que habla la ciencia física moderna y aquello de lo que no habla. Si uno delimita con cuidado «ciencia física» de modo que incluya la óptica pero no llegue hasta la psicofisiología, entonces puede decir que la ciencia habla de la dureza y de las longitudes de onda de la luz, pero no de la marronez o de la fealdad. Hasta ahí, todo bien. Pero no debemos decir a continuación que aquello de lo que la ciencia física no habla es automáticamente «subjetivo». Si lo hacemos, parecemos estar diciendo que no hay ningún modo verdaderamente racional de zanjar los debates sobre cualquier cosa de la que la ciencia no se ocupe. Pero eso parece obviamente falso.

Creo que nuestro uso de nociones borrosas como «subjetivo» y «objetivo» es simplemente un caso de retardo cultural. Todavía estamos bajo los efectos de la revolución científica del siglo XVII, el periodo en que la cosmología a escala humana de la *Divina comedia* de Dante fue sustituida por la aterradora cosmología de Galileo y Laplace, el universo en el que todo lo que ocurre se puede predecir conociendo algunas leyes simples y las posiciones de las partículas elementales. Locke estaba tan impresionado por esa nueva ciencia, que incrustó en la filosofía moderna la distinción de cualidades primarias-secundarias formulada como la distinción entre estar realmente ahí y ser solo una de nuestras «ideas». Pero ya no deberíamos seguir impresionados. Hemos convivido el tiempo suficiente con la ciencia física moderna para no verla como una revelación sobre la condición humana, sino simplemente como un modo nuevo y mejor de averiguar cómo funcionan las cosas y predecir qué va a ocurrir a continuación.

Esto se puede expresar de otra manera considerando la relación entre ciencia y religión. El dios al que los cristianos adoraban antes de Galileo presidía un universo finito, figurable, hogareño. A despecho del discurso teológico profesional sobre que «está en todas partes y en ninguna», el común y corriente cristiano de a pie sabía que Dios ocupaba una cierta posición dentro del universo: arriba del todo, por decirlo así. Sin embargo, una vez que el universo se describió como infinito y mecánico, esa imagen saltó en pedazos. Así, abandonar la esperanza, el consuelo de la creencia religiosa, se convirtió en la marca del intelectual. Los intelectuales se preciaron de aceptar un universo de átomos que iban por ahí cargándose ciegamente. Pero, en su ansia por mostrarse libres de superstición, se pasaron de la raya. Crearon una mala distinción filosófica, entre aquello de lo que habla la ciencia y todo lo demás, diciendo que aquello de lo que no habla contaba como «subjetivo», o «relativo», o «cultural», o «en el ojo del observador».

Esa distinción simplona debería abandonarse sin más. Deberíamos sustituirla por una escala deslizante de disputabilidad. En un extremo están la lógica y las matemáticas, donde un enunciado es demostrable o no lo es. Un poquito más adelante en la

escala están las cuestiones empíricas indisputables, como la de si un determinado árbol es un olmo o un haya, si un determinado compuesto contiene carbono, si una determinada mesa es marrón, si una determinada dolencia es sarampión. Más adelante todavía están cuestiones más controvertidas como la de si determinada mesa es fea, si determinado matrimonio se puede salvar, si determinado político debería ser elegido. En el extremo más lejano de la escala están disputas como la que hay entre marxistas y liberales sobre la forma de escribir la historia, entre la concepción cristiana de la divinidad y la budista, entre la concepción aristotélica y la galileana del universo, entre la concepción de Kant y la de Nietzsche sobre cómo vivir tu vida, entre el tipo de pintura que toma a David como paradigma y la que toma como paradigma a Cézanne.

Esa escala que va desde el álgebra en un extremo hasta el arte en el otro parece preferible a la distinción objetivo-subjetivo, porque no depende de la pintoresca noción de «estar en la mente» en cuanto que opuesta a «estar ahí afuera», ni de sobrestimar equivocadamente la significación de la ciencia física. En la concepción que yo quiero recomendar, el grado de disputabilidad de una cuestión no está en función del carácter metafísico de su tema. Ni tampoco de una distinción entre facultades humanas —por ejemplo, la «razón» como opuesta al «gusto»— ni entre métodos de investigación —por ejemplo, la «lógica» como opuesta a la «observación»—. Todas esas distinciones se me antojan demasiado simples como para iluminar nada. Si queremos entender por qué los físicos a menudo están de acuerdo y los teólogos a menudo no, o por qué sí los matemáticos y no los críticos literarios, lo que nos ayudará no son distinciones filosóficas tradicionales como «objetivo-subjetivo», o «racional vs. irracional», o «conceptual vs. empírico». Lo único que nos ayudará es fijarnos en cómo se las arregla de verdad la gente para resolver las disputas en esos diversos campos.

III
Los argumentos de Mackie en favor de la «subjetividad» de los valores

Para explicar con más claridad por qué creo que «subjetivo» y «objetivo» son términos inservibles, quiero criticar algunos argumentos que pretenden demostrar que los valores son subjetivos. En el primer capítulo de *Ética: la invención de lo bueno y lo malo*, el reciente libro de J. L. Mackie, el autor cita a otro filósofo de Oxford —Richard Hare— cuando dice que este último «no entiende qué quiere decirse con "la objetividad de los valores"» y que «no ha conocido a nadie que lo entienda». Hare había argumentado de la siguiente manera:

> Piensa en un mundo en cuyo tejido haya valores objetivamente inscritos; y piensa en otro donde esos valores hayan sido aniquilados. Y recuerda que, en ambos, la gente que hay en ellos sigue preocupada por las mismas cosas; no hay diferencias en la preocupación «subjetiva» que tienen por las cosas, solo la hay en su valor «objetivo». Ahora pregunto: «¿qué diferencia hay entre los estados de cosas de esos dos mundos?» ¿Puede darse otra respuesta que no sea «ninguna en absoluto»?[2]

Creo que Hare tiene toda la razón en esto. El mismo argumento, pienso yo, valdría para demostrar que la afirmación «los colores no son propiedades objetivas de los objetos materiales» es una tontería; supongamos que hubiera un mundo en que lo fueran y otro en el que solo se lo pareciera a la gente, siempre y en todas partes. ¿Qué demonios nos va a importar en cuál de ellos estamos viviendo? Como dijo William James, toda diferencia debe marcar alguna diferencia. El verdadero problema de la distinción subjetivo-objetivo, ya se aplique a los colores o a la belleza o a la bondad, es precisamente que el que algo sea subjetivo u objetivo no marca diferencia alguna en nuestra vida. Discutir esa cuestión, como James dijo, es como discutir si son gnomos o trasgos los que mueven los hilos del mundo. Lo que marca una diferencia es si disponemos o no de procedimientos para resolver disputas sobre colores o sobre valores. Pero eso es una cuestión de nuestras prácticas, no de «lo que está ahí en el mundo».

Mackie replica al argumento de Hare del modo siguiente:

> El argumento de Hare es similar a la tesis positivista de que no hay diferencia entre un mundo fenomenista o berkeleyano, donde solo hay mentes y sus ideas, y el mundo realista de sentido común en el que también hay cosas materiales, pues es lógicamente posible que la gente tuviera las mismas experiencias en ambos. Si rechazamos el positivismo que convertiría en una pseudo-cuestión la disputa entre realistas y fenomenistas, podemos rechazar la negativa de Hare a considerar el problema de la objetividad de los valores, que se apoya en algo parecido.[3]

Con «positivismo», Mackie quiere decir la tesis de que las cuestiones que no sabemos cómo decidir no son cuestiones reales, «no tienen significado». Los positivistas lógicos decían que el debate entre fenomenistas y realistas en torno a si realmente hay objetos materiales era «verbal» porque ninguna evidencia sensorial ayudaría a decidirlo, ya que ambas partes estaban de acuerdo en que los sentidos nos darían exactamente los mismos informes independientemente de quién tuviera razón. A diferencia de Mackie, yo creo que aquí los positivistas iban bien encaminados. Quizá no deberíamos decir que el debate es «verbal» o «sin significado», pero podemos decir razonablemente que, a menos que a alguien se le ocurra un modo de discutirlo, haríamos bien en olvidarnos de él. Así que parece justo preguntarle a Mackie cómo cree que podría discutirse la objetividad de los valores. Él presenta tres argumentos para afirmar que no son objetivos, y me ocuparé brevemente de cada uno de ellos.

El primero es «a partir de la relatividad» y dice que

> las variaciones reales entre códigos morales se explican más fácilmente con la hipótesis de que reflejan modos de vida que con la de que expresan percepciones, en su mayoría seriamente inadecuadas o distorsionadas, de valores objetivos.[4]

¿Qué es explicar una diferencia de código moral como el reflejo de un modo de vida? Seguramente lo que Mackie tiene en mente son ejemplos como la afirmación de que en ciertas tribus hay determinados tabúes debido a sus necesidades socioeconómicas.

Pero ese tipo de explicación también funciona bastante bien con la ciencia: los pre-galileanos aceptaban una cosmología aristotélica en parte porque no tenían la tecnología que les habría permitido fabricar lentes telescópicas. La ciencia galileana no habría podido ser ampliamente difundida y aceptada de no haber sido por la Reforma. Y así sucesivamente. Para probar su tesis, Mackie tendría que demostrar que las explicaciones mediante factores «extrínsecos» son más apropiadas en el caso de códigos morales diferentes que en el de creencias científicas diferentes. Su argumento a partir de la relatividad pivota sobre la idea de que explicamos la elección entre teorías científicas en términos de una mejor o peor percepción de los objetos externos, mientras que no hacemos igual al elegir opiniones morales. Pero esa es una concepción empirista simplona de cómo funciona la ciencia. Es la explicación de la ciencia que *La estructura de las revoluciones científicas* de Kuhn nos ha ayudado a dejar de lado. Mackie está retratando al físico como si estuviera constreñido a captar correctamente la estructura de la materia a base de encuentros con esa estructura, pero se niega a ver al moralista como alguien que capta correctamente la obligación moral a base de encuentros con la ley moral. Uno puede hacer esa distinción, pero solo si previamente está persuadido de que las estructuras atómicas están ahí afuera para constreñir nuestras creencias y que la ley moral no lo está.

Sin esa presuposición, uno no podría emplear un «argumento desde la relatividad». Pero la presuposición solo puede establecerse con algo parecido a lo que Mackie llama un «argumento desde la rareza». De hecho, ese es su segundo argumento en contra de la objetividad de los valores. Mackie dice que tiene dos partes, una metafísica y otra epistemológica. La parte metafísica dice que «si hubiera valores objetivos, serían entidades, o cualidades, o relaciones, de un tipo muy extraño, completamente diferentes a cualquier otra cosa del universo».[5] La parte epistemológica dice que «si nos percatáramos de ellos, tendría que ser mediante una facultad especial de percepción o intuición moral, completamente diferente de las formas que tenemos de conocer todo lo demás».[6] La razón

por la que los valores tendrían que ser algo completamente diferente, afirma Mackie, es que «intrínsecamente guían o motivan acciones», pero no conocemos ningún otro tipo de objetos que hagan eso. Además, dice, es un problema «cómo podrían dichos valores resultar o sobrevenir a partir de rasgos naturales», con lo cual se refiere a que tendríamos que explicar cómo encajan en la descripción científica del mundo, qué lugar puede dejar la física para ellos.[7]

Yo creo que, si uno acepta el principio de que nada que no figure en una descripción científica del mundo puede ser real, como hacen Locke y Mackie, es verdad que los valores le resultarán raros. Pero, salvo en ese caso, que haya entidades en el mundo que intrínsecamente guíen o motiven acciones no parecerá extraño. Si pensamos en una potencial pareja sexual o en un trozo de comida como el sentido común piensa en ellos, entonces parecerá natural decir que guían o motivan intrínsecamente acciones. Si los pensamos como una simple reunión de átomos «realmente», entonces no. Análogamente, pensamos que el espectáculo de la vida humana heroica —la de Sócrates, o la de San Francisco de Asís, o la de Sájarov— guía acciones intrínsecamente, que uno solo podría dejar de motivarse con vidas así si sus facultades estuvieran disminuidas de alguna manera. Pero, por supuesto, hay montones de vocabularios en los que podrían contarse sus historias —por ejemplo, un vocabulario fisicista en términos de las partículas que las componen, uno conductista, uno psicoanalítico— que no sugerirían que tales vidas deban imitarse. Mackie piensa que es raro que algo sea guía de acciones «intrínsecamente» porque piensa que cualquier descripción de ese objeto que es guía de acciones mete de contrabando alguna propiedad extrínseca, meramente relacional. Pero solo puede pensar eso si, como Locke, da por hecho que cualquier vocabulario que no sea el de las ciencias físicas introduce ese tipo de propiedades. Una vez más, me parece que su argumento se reduce a afirmar que el vocabulario de la ciencia física no es el vocabulario de la elección moral. Uno solo tendrá la impresión de que hay un problema con la relación entre valores y cualidades

«naturales» si previamente ha definido las cualidades «naturales» de tal manera que solo cuenten las que nombra el vocabulario de la ciencia física.

Pasando ahora al lado epistemológico del argumento de Mackie desde la rareza, dice él que le parece ad hoc postular una facultad específica de intuición moral que reconoce los valores de objetos o sucesos. Así es, pero no parece ad hoc decir que nuestra capacidad de alcanzar acuerdos racionales sobre cuestiones morales se parece mucho a nuestra capacidad de alcanzar acuerdos racionales sobre teorías científicas. Volvemos otra vez al empirismo más bien simplista de Mackie. Él piensa que todo conocimiento, o bien tiene que proceder de los cinco sentidos, o tiene que basarse de manera clara en un conocimiento que proceda de ahí. Piensa que no hay ningún argumento del «es» al «debe» porque cree que todas las premisas-«es» deben reducirse en última instancia a premisas sobre lo que alguien ve, u oye, o palpa. Pero esta clase de empirismo no puede darle más sentido a la elección de teorías en ciencia del que puede darle a la elección de teorías en moral. Si de verdad pensáramos que solo tenemos «conocimiento objetivo» cuando podemos maniatar la elección de teorías a la percepción sensorial, deberíamos concluir que prácticamente no tenemos conocimiento objetivo. Mackie no está dispuesto a admitir ninguna explicación de nuestro conocimiento objetivo de valores salvo en la forma de un sexto sentido, de alguna facultad modelada sobre la de la vista. Pero semejante limitación de la explicación epistemológica da por supuesto que saber de valores es muy parecido a saber de formas o tamaños. Mas ¿por qué habría de ser parecido? ¿Por qué llegar a entender qué tipo de vida es el mejor para un ser humano tendría que ser como ver lo grande que es una cosa?

El argumento tercero y último de Mackie contra la objetividad de los valores es que resulta más fácil explicar las creencias morales de la gente apelando a sus deseos, sus relaciones de poder dentro de la sociedad, sus esperanzas y miedos, y cosas parecidas, que apelando a hechos morales de los que se llega a tener conocimiento. Este argumento lo repite de modo sucinto mi colega Gilbert Harman en

su libro *La naturaleza de la moralidad* cuando dice que la gran diferencia entre la ciencia y la ética es que

> para explicar que ocurran las observaciones que sustentan una teoría científica necesitas hacer suposiciones sobre determinados hechos físicos, pero no parece que las hagas sobre ningún hecho moral para explicar que ocurran las así llamadas observaciones morales.[8]

El ejemplo que brinda Harman de una «así llamada observación moral» es «doblar la esquina y ver a un grupo de jóvenes matones echando gasolina a un gato y prendiéndole fuego». Aquí estamos tentados de decir que vemos sin más que una acción está mal. Pero, pregunta Harman, «¿nuestra reacción se debe a la maldad real de lo que ves, o es simplemente una reflexión de nuestro "sentido" moral, un "sentido" que has adquirido tal vez como resultado de tu educación moral?».[9] Él se inclina a pensar que lo que mejor explica nuestra reacción es siempre lo segundo, y Mackie coincide. Mackie apunta a las mismas consideraciones que utilizó para apoyar su argumento desde la relatividad: la capacidad de psicólogos, antropólogos e historiadores de explicar, sin mencionar en ningún momento hechos morales, por qué la gente afirma ver que esto y aquello está bien o está mal, es bueno o es malo.

Para valorar este punto, necesitamos considerar si lo que mejor explica que alguien haga una observación en casos no morales es la presencia de lo que la persona afirma observar. Dice Harman que, en el caso de un científico que está vigilando una cámara de niebla, ve un rastro de vapor y exclama «¡ahí va un protón!», la mejor explicación de su observación incluye el hecho de que había efectivamente un protón atravesando la cámara. En el caso de una persona que les dice «eso está mal» a los matones que torturan al gato, en cambio, afirma que «parecería que todo lo que necesitamos suponer es que tiene determinados principios morales más o menos articulados […] Si su juicio intuitivo inmediato es verdadero o falso parece completamente irrelevante para nuestra explicación».[10] Creo que Harman tiene razón en esto último, pero que se equivoca en el ejemplo del protón. A mí me parece que, también ahí, es

completamente irrelevante para explicar la observación de los científicos el que haya o no protones. Supongamos que los físicos nucleares empiezan a decirnos que nunca hubo protones, que el concepto mismo de «protón» era una forma engañosa de describir la estructura de la materia, y que no hay ningún modo claro de emparejar tipos de rastros de vapor en cámaras de niebla con tipos de partículas elementales. De hecho, supongamos que la noción misma de «partícula elemental» pasa a parecer un modo primitivo y desorientador de hablar sobre discontinuidades en el espacio-tiempo, o sobre algún otro tipo remoto y novedoso de entidad. Todos esos cambios parecen dejar intacta nuestra explicación de la observación de un protón. Sabemos cómo se formó al investigador para hacer informes sobre protones, y sabemos que hizo lo que se le había dicho; así que sabemos por qué informó de lo que informó. La cuestión de si lo formaron para usar los términos de una teoría física falsa sencillamente no se plantea. Análogamente, podemos explicar por qué tenemos una reacción negativa a la tortura de gatos porque hemos sido criados en una atmósfera cristiana, sentimental y burguesa. Esa explicación es del todo suficiente, y sería irrelevante plantear la cuestión ulterior de si las opiniones sentimentales sobre la bondad para con los animales captan o no correctamente los hechos morales. Pero eso no significa que no sea tan objetivamente bueno o malo torturar gatos, como objetivamente correcto o incorrecto es decir que hay protones.

Lo que quiero objetar aquí a Harman y Mackie es, de nuevo, que su intento de distinguir entre ciencia y ética se apoya en un enfoque empirista simplón de la verdad científica. Mackie cree que las teorías científicas se pueden reducir de algún modo a informes de observaciones sensoriales. Harman cree que «para explicar que ocurran las observaciones que sustentan una teoría científica necesitas hacer suposiciones sobre determinados hechos físicos». Ninguna de las dos opiniones me parece plausible. Las teorías científicas, como han dicho Quine y Kuhn, tocan nuestra experiencia solo intermitente e incidentalmente. Eso lo tienen en común con las perspectivas morales. Ni desarrollamos las teorías científicas a base de intentar dar cuenta de la apariencia sensorial de las cosas, ni

desarrollamos las teorías morales a base de intentar dar cuenta de nuestras reacciones morales a bote pronto, aunque ocasionalmente podamos hacer ambas cosas.

Puedo concretar más esta crítica citando un último pasaje de Mackie. Él resume su visión sobre lo que es bueno para el hombre, o sobre la meta de la vida humana, diciendo:

> El razonamiento moral consiste en parte en lograr una comprensión más adecuada de esta meta básica (o conjunto de metas), y en parte en dar con el mejor modo de perseguirla y realizarla. Pero este enfoque está abierto a dos interpretaciones radicalmente distintas. Según una de ellas, decir que algo es lo bueno para el hombre o la meta general de la vida humana no es más que decir que es eso lo que los hombres de hecho persiguen o lo que encontrarán satisfactorio de modo último, o quizá algo que, si se postula como una meta implícita, nos permite dar sentido a lo que efectivamente son los esfuerzos humanos y detectar un patrón coherente en lo que de otro modo parecería un revoltijo caótico de propósitos en conflicto. De acuerdo con la otra interpretación, decir que algo es lo bueno para el hombre o la meta general de la vida humana es decir que ese es el fin propio del hombre, que eso es lo que de hecho *debe* esforzarse por perseguir, lo haga o no de hecho.[11]

En mi opinión solo hay una diferencia estilística entre esas dos interpretaciones. Si encontramos la forma de detectar un patrón coherente en lo que de otro modo parecería un revoltijo de propósitos en conflicto, y si actuar conforme a ese patrón es satisfactorio de modo último, ¿qué se añade diciendo que ese es de hecho «el fin propio del hombre»? Mackie piensa que la diferencia entre las dos interpretaciones es la diferencia entre hacer un enunciado descriptivo y uno prescriptivo, entre decir algo sobre psicología y hacer una afirmación sobre valores objetivos. Yo pienso que la diferencia es como la que hay entre decir «esta teoría del universo físico parece dar cuenta de todos los fenómenos mejor que cualquiera de sus competidoras, y la hemos utilizado para conseguir un dominio extraordinario sobre la naturaleza» y decir «esta teoría nos dice cómo funciona efectivamente el mundo físico». No hay forma de enemistar a un enunciado con el otro, hasta donde yo

alcanzo a ver. No podemos tener buenas razones para decir que una teoría funcionará maravillosamente para siempre, en todos los campos, pero que aún así no nos dice realmente nada sobre el mundo. Análogamente, no veo forma de enfrentar una con otra las dos interpretaciones de Mackie de manera que tuviera sentido decir que una teoría moral cumple todos los fines que un sistema ético haya podido esperar cumplir jamás, pero que desgraciadamente es meramente «subjetiva». Puedo ver diferencia entre decir, como cuestión de hecho psicológica, que la mayoría de la gente disfruta torturando gatos, y decir, prescriptivamente, que deben disfrutar con eso. Pero, cuando la cuestión es acerca de una teoría global sobre lo correcto y lo incorrecto —una que es satisfactoria de modo último—, no veo cómo o por qué haría uno esa distinción.

IV
Principios morales

Esto es todo lo que diré sobre el intento de Mackie de darle sentido a la idea de que los valores son meramente subjetivos. Pero quiero recalcar un poco más mi tesis sobre las analogías entre ciencia y ética abordando el tema de los principios morales. Los filósofos en la línea de Mackie dicen a menudo que es imposible que un enunciado como «uno debe reprimir sus impulsos de ser cruel con los gatos» sea verdadero del modo en que lo es «los cuerpos se atraen entre sí con una fuerza proporcional al cuadrado de su distancia». La idea es que los enunciados-«debe» no pretenden decirnos cómo son los objetos, cosa que sí hacen los enunciados sobre la gravedad, por ejemplo. Para esta opinión, buscar valores objetivos —objetos de los que sean verdaderos los enunciados-«debe»— es sencillamente un error porque los enunciados-«debe» no son susceptibles de ser ni verdaderos ni falsos. En mi propia opinión, los enunciados-«debe» sobre la crueldad son verdaderos exactamente de la misma manera que los enunciados-«es» sobre la gravedad. Ni unos ni otros son verdaderos de objetos. En ninguno de los dos casos tiene sentido alguno pensar en la verdad como «correspondencia con la realidad». No hay

ningún objeto llamado «gravedad» con el que corresponderse, como no hay ninguno que respetar llamado «lo debido». El sentido y la importancia de ambos enunciados está en ser extractos de visiones a gran escala sobre cómo debemos vivir o cómo debemos entender lo que está pasando. No pueden emparejarse con la realidad punto por punto, ni se gana nada pensando que «informan de lo que realmente hay ahí afuera». Ambos se parecen a enunciados como «hay números tales que su cuadrado es igual al propio número». Eso es parte de una teoría y se acepta como verdadero porque la teoría como un todo cumple las necesidades para las que se construyen teorías sobre los números, pero no hay forma de confrontarlo con objetos. Aun así, sería una tontería decir que es «meramente subjetivo», o «meramente relativo», o «meramente convencional».

La principal razón por la que la gente dice que los enunciados-«debe» no pueden ser verdaderos en el mismo sentido en que lo son las teorías científicas es que parece relativamente fácil imaginar teorías de la moralidad rivales basadas en primeros principios contrapuestos. Supongamos que una teoría parte de la premisa de que uno debe expresar sus impulsos sádicos, y otra de que debe reprimirlos. Dada esa oposición, se nos dice, no habrá forma de resolver la disputa entre las dos teorías, que son igualmente coherentes y defendibles. Pero a mí esta afirmación me parece palmariamente falsa. La entera idea de una teoría moral que empieza por alentar la crueldad no es más que una ficción del filósofo; no hay tales teorías, como no hay teorías matemáticas que empiecen con la proposición de que no hay infinitos números o de que la división larga no funciona. No puedes probar la relatividad o la subjetividad de la moral diciendo que alguien podría sostener con terquedad y coherencia una opinión estúpida, más de lo que puedes probar la relatividad o la subjetividad de la ciencia física diciendo que alguien podría afirmar coherentemente que la Tierra es plana o que lo que Newton llamaba «gravedad» es obra en realidad de unos duendes invisibles. Una teoría científica, para ser tomada en serio, no debe negar sin más los primeros principios de otra, debe darnos a continuación su propia explicación de cómo funciona algo y

demostrar las ventajas de esa explicación. De manera parecida, para que una visión de la moralidad pueda tomarse en serio, no puede proclamar sin más que la crueldad y el egoísmo son buenos; a continuación debe desarrollar una visión de la vida humana y de la sociedad que incluya esas afirmaciones, les dé sentido y demuestre sus ventajas. Eso no es fácil de hacer, no más fácil que explicar que la Tierra solo parece redonda pero en realidad es plana.

La misma idea se puede expresar afirmando que no hay forma de refutar a un hombre que dijera que su posición moral es que todo el mundo debe sacrificarse a él, porque él es más importante que nadie; o, como Hitler, que nadie tiene derechos que haya que respetar excepto los alemanes de derechas. Es cierto que no hay modo de refutar a gente así, pero tampoco hay modo de refutar a quien dice que no existen números por encima de un millón, o que el mundo está gobernado por un comité de trasgos y gnomos, o que todos menos él somos robots. Hay montones de cosas idiotas, o absurdas, o delirantes, que se pueden decir y no pueden refutarse. Ello es así porque las nociones de argumento, desacuerdo, refutación, controversia, y similares, solo tienen sentido en situaciones donde ambos lados comparten un cierto acuerdo mínimo. Solo se puede refutar a alguien que comparte algunas de las premisas propias. A quienes comparten un número demasiado bajo de premisas no se los trata como oponentes irrefutables, sino como demasiado pueriles, o chiflados, o iletrados, o lo que sea, como para entrar en la conversación. Toda esfera de discurso —moral, científico, matemático, literario— tiene, por así decir, determinados requisitos de admisión, determinadas credenciales que hay que exhibir antes de ser tomado en serio. La diferencia entre Hitler y, digamos, Nietzsche, es que Hitler carece de las credenciales y Nietzsche las tiene. Nietzsche comparte las preocupaciones, los intereses, el vocabulario, de Aristóteles y Mill y Kant. Hitler no. A Nietzsche se lo puede discutir, aunque quizá no «refutar» en el sentido en que un matemático o un jurista pueden refutar a otro. A Hitler con toda seguridad no se lo puede refutar. Pero no porque tenga una posición irrefutable, sino sencillamente porque no se lo puede discutir en absoluto.

La diferencia entre Nietzsche y Hitler es paralela a la que hay entre Einstein y la Sociedad Terraplanista. Einstein dijo algunas cosas chocantes, paradójicas, absurdas casi, pero las dijo en el contexto de la teoría física de su tiempo, en respuesta a problemas dentro de esa teoría. Consiguió transformar la física demostrando que podías conseguir explicaciones mejores aceptando las paradojas. Los de la Sociedad Terraplanista, en cambio, ni saben lo que se dice dentro de la teoría científica ni les importa; lo único que quieren es demostrar que nadie puede probar que la Tierra sea redonda. En eso llevan toda la razón, igual que la Sociedad para la Crueldad con los Gatos tendría razón en mantener que nadie puede probar que no debas torturar gatos. Pero eso no es decir que unos u otros tengan mucho que aportar a la discusión racional de sus respectivos temas.

V
Filosofía, literatura y ciencia social

Creo que una de las razones de que el problema de la objetividad de los valores parezca más interesante y serio de lo que es, es que la gente piensa que el conocimiento moral es la provincia de un grupo de especialistas: los filósofos morales. A algunos filósofos les gusta trasladar la impresión de que, si realmente quieres oír la verdad sobre cómo vivir tu vida u organizar tu sociedad, lo natural es que te vuelvas hacia Platón, Kant, Nietzsche, Rawls y gente así. A mí esa impresión me parece terriblemente engañosa. Durante los siglos XVIII y XIX, cuando la religión estaba perdiendo su prestigio como fuente de consejo e iluminación, la filosofía se ofreció como recambio. Aquella fue la gran era de la filosofía moral. En ese periodo, Kant y Bentham propusieron fórmulas bonitas, simples, abstractas, que resumían y clarificaban supuestamente todas nuestras intuiciones morales. Supuestamente, incluso nos permitían resolver angustiosos dilemas morales con la aplicación limpia y directa del razonamiento inductivo o deductivo. Pero casi en el mismo instante en que el imperativo categórico y el principio utilitario quedaron lanzados, la gente se dio cuenta de que eran tan

inútiles como los Diez Mandamientos o la Regla de Oro. Sea lo que sea el conocimiento moral, no se alcanza memorizando reglas o formulando principios. Lo último que uno quiere cuando se enfrenta a un dilema ético o político genuino son principios generales.

Esta observación puede sonar chocante. A menudo describimos a una persona moralmente mala como «sin principios», y elogiamos a las que son buenas por tener principios «firmes» o «elevados». Así que formular principios puede parecer tremendamente importante. Pero lo que realmente queremos decir con esas expresiones es solo que hay cosas que una buena persona sencillamente no haría en ninguna circunstancia: no traicionará a sus amigos, no estafará a sus clientes, no defraudará la confianza. Alguien de elevados principios no es necesariamente alguien que puede formular un principio general y deducir de él acciones particulares. Describirlo como dotado de principios firmes o elevados se parece mucho más a decir que tiene un carácter recto o una gran dignidad moral. No es una observación sobre lo que sabe, sino sobre lo que cabe esperar que haga en un aprieto. Si de verdad conociéramos a alguien que intentara deducir de principios generales lo que debe hacer, creo que recelaríamos un poco de él. Porque la idea de que los dilemas morales se pueden resolver de esa manera tiene algo de extraño; tanto como la de que uno puede elegir un amigo o un amante haciendo una lista de los valores que él o ella debería ejemplificar y buscando luego a alguien que los maximice.

En el siglo XX nos vemos como moralmente más avanzados que generaciones anteriores, que nuestras elecciones éticas son más sofisticadas y lúcidas y sabias. Pero no es porque creamos conocer más verdades morales generales que ellas. Es más bien porque tenemos más experiencia, en el sentido de tener un sentido más vívido de las alternativas, de las consecuencias que se pueden seguir de diferentes elecciones. En particular, no poseemos esa imaginación más vívida y esa cultura más amplia porque tengamos sistemas más sofisticados de filosofía moral, o sea, principios generales más convincentes o escalas de valores más claramente definidas. Es porque somos beneficiarios de dos desarrollos culturales que en tiempos de Kant y Bentham estaban solo iniciándose. Se trata, dicho

a grandes rasgos, de la literatura moderna y las ciencias sociales. En tiempos de Kant ni se soñaba todavía con novelas como las de Dostoievski y Dickens, Balzac y James, Kafka y Proust, Nabokov y Lessing. La reflexión sobre el curso de la vida humana y los modos posibles en que puede vivirse todavía estaba por encajarse con ideas que dejaron el estoicismo y el cristianismo; las formas increíblemente variadas de describirse a uno mismo que debemos a esos novelistas sencillamente no estaban disponibles aún. Ni siquiera Goethe, cuya imaginación era tan vasta que dejaba anonadados a sus contemporáneos, pudo entrever ni una fracción de los tipos de vida humana que iban a proponerse como alternativas en los cien años siguientes a su muerte. En nuestro siglo es en las novelas, no en los sermones o en los libros de filosofía moral, donde encontramos los vocabularios que usamos para formular decisiones morales. Nos preguntamos si, al abrazar este o aquel plan de vida, no correremos el riesgo de ser como determinado personaje; a veces un personaje histórico, pero más a menudo el personaje de alguna novela. Cuando intentamos encontrar el modo de formular por qué detestamos a ciertas personas y admiramos a otras, es su comparación con personas de la literatura lo que proporciona el método más fácil para hacernos claros nuestros criterios a nosotros mismos. En todo lo que tiene que ver con la ética privada más que con la pública, ser moralmente sofisticado no es tener opiniones filosóficas avanzadas, sino tener una imaginación educada por las obras literarias.

En lo que se refiere a las decisiones públicas más que a las privadas —a la ética social por contraposición a la ética personal—, nuestra sofisticación no nos viene de la reflexión filosófica sobre la naturaleza de la justicia social, sino de la antropología, la psiquiatría, la historia política y cultural, la política comparada y la sociología. Nuestro siglo se ha beneficiado de la extraordinaria proliferación de intentos por comprender culturas extrañas, formas extrañas de personalidad, extrañas formas de organizar la vida política. Gracias a genios como Marx y Freud, pero también gracias a la pura acumulación de datos antropológicos e históricos, hemos adquirido una visión mucho más tolerante y con bases mucho más amplias de

las posibilidades que tienen abiertas los seres humanos y las sociedades. Ese incremento de la tolerancia, la capacidad aumentada de ver a otras clases sociales, o grupos, o culturas, o sistemas políticos, como extraños pero no necesariamente inmorales, es una buena parte de lo que queremos decir cuando decimos que somos moralmente más sofisticados que nuestros predecesores del pasado siglo.

Déjenme ilustrar mi tesis sobre el papel de la literatura y las ciencias sociales discutiendo la siguiente pregunta: dado que el enorme cambio habido en Estados Unidos en los últimos treinta años en el trato que los blancos dispensan a los negros es un avance moral, ¿sobre la base de qué conocimiento se hizo ese avance? El movimiento por los derechos civiles fue primordialmente un asunto de presión moral más que de fuerza. Al fin y al cabo, los blancos controlaban la policía y el ejército y podían haber eliminado salvajemente el movimiento metiendo a sus líderes en campos de concentración y aumentando el rigor de la legislación discriminatoria. Pero no lo hicieron. ¿Fue porque los filósofos habían hallado argumentos nuevos en favor de un principio moral —a saber, que el color de la piel no cuenta—, principio que antes era dudoso? ¿O porque habían aprendido algo sobre la clasificación de diversos valores objetivos que antes ignoraban? Dudo que ninguna de esas explicaciones sirva. Comprenderemos mejor el fenómeno si recordamos que la generación de clase media blanca que gobernaba Norteamérica en las décadas de 1950 y 1960 había crecido leyendo novelas como *Chico negro*, de Richard Wright, y tratados sociológicos como *An American Dilemma*, de Gunnar Myrdal. Se les había puesto delante de las narices qué era ser negro en Estados Unidos y se les había hecho sentirse culpables. Cuando apareció el movimiento por los derechos civiles, respondieron en consecuencia. Esto es un ejemplo de lo que quiero decir cuando digo que nuestros educadores morales son la literatura y las ciencias sociales, no la filosofía. Los intentos de los filósofos por formular principios y clasificar valores no son inútiles, pero solo son útiles si otras áreas de la cultura nos han hecho moralmente más sensibles.

Mi conclusión de todo esto es que sería de ayuda que arrumbáramos, no solo la distinción entre subjetivo y objetivo, sino entre hechos y valores, entre descripción empírica y valoración del estado de cosas descrito. Esas distinciones, antes que facilitar la reflexión moral, la han hecho más difícil. Como he sido más bien duro con la filosofía como disciplina en este artículo, permítanme que termine tratando de corregir el equilibrio. La mayor parte de lo que he estado diciendo está tomado de Wittgenstein, Dewey y Heidegger, que me parecen los grandes filósofos de nuestro siglo. Su intento de romper dualismos filosóficos tradicionales como los que he venido criticando y ponernos de nuevo en contacto con la deliberación moral real me parece demostrar que la única cura para la mala filosofía es más filosofía. Así que, a pesar de todas las cosas feas que he dicho de las distinciones filosóficas tradicionales, pienso que la filosofía todavía tiene una función social. No es descubrir hechos inesperados ni demostrar nada. Es más bien proponer un vocabulario, un modo de enunciar nuestras preocupaciones, que nos ayude a dar sentido a lo que hacemos, a ser más conscientes de que podríamos estar equivocados y a aumentar nuestra sensibilidad a nuevas posibilidades.

---

[1] N. del T.: «severa» y «blando» traducen aquí, respectivamente, «*tough-minded*» y «*tender-minded*», los términos (de resonancias múltiples) con que William James caracterizaba los dos grandes «temperamentos filosóficos» en la Conferencia I de *Pragmatismo*. Ramón del Castillo usa para ellos los adjetivos «rudo» y «selecto» en su cuidada traducción del libro de James (Madrid, Alianza Ed., 2016); no seguimos ese uso en esta ocasión por razones simplemente estilísticas, pero conviene dejar anotada esta primera alusión de Rorty a James en el artículo.

[2] J. L. Mackie, *Ethics: Inventing Right and Wrong*. Nueva York: Pelican Books, 1977, p. 21 [hay trad. esp.: *Ética: la invención de lo bueno y lo malo*. Barcelona, Gedisa, 2008].

[3] *Ibid*., p. 22.

[4] *Ibid*., p. 37.

[5] *Ibid*., p. 38.

[6] *Ibid*.

[7] *Ibid*., p. 49.

[8] Gilbert Harman, *The Nature of Morality: An Introduction to Ethics*. Nueva York: Oxford University Press, 1977, p. 6 [hay trad. esp.: *La naturaleza de la moralidad. Una introducción a la ética*, México, UNAM, 1998].

[9] *Ibid*., p. 4.

[10] *Ibid*., p. 7.

[11] Mackie, *Ethics*, pp. 46-47; las cursivas son mías.

CAPÍTULO 12
QUÉ ESTÁ MUERTO DE PLATÓN

Es muy cierto que la historia de la filosofía occidental es una serie de notas a pie de página a Platón, pero, de unos pocos siglos a esta parte, muchas de esas notas han estado diciendo: «obsérvese cuánto daño ha hecho esta mala idea platónica en concreto». Dante Germino tiene toda la razón en que Platón ha sido frecuentemente simplificado y caricaturizado. Pero también han discrepado de él, sobre asuntos fundamentales, personas que pueden arrogarse buena comprensión de su obra. Estoy de acuerdo con que el Platón de Karl Popper es una caricatura. Pero creo que filósofos como Nietzsche, Dewey, Wittgenstein y Heidegger no lo caricaturizaron y nos han dado buenas razones para descartar la mayoría de sus doctrinas centrales.

Esos filósofos han creído, entiendo que a diferencia de Germino, que Platón ofreció un «sistema intelectual». Puede que, como Germino sugiere, el sistema que han hallado los comentaristas en los libros centrales de *La república* y en diálogos como *Fedón* y *Teeteto* no representen la verdadera opinión de Platón, que ese sistema sea una suerte de elaborada broma o un dispositivo pedagógico pensado para conducir más allá de él mismo. De hecho, Platón dice en una de sus cartas que su verdadera doctrina no puede ponerse por escrito. Los comentaristas no han sabido bien qué hacer con esa carta, y yo tampoco. Así que el «Platón» que voy a discutir es el autor del sistema intelectual conocido como platonismo, el que contiene doctrinas como la de las Formas y que la integra con las distinciones entre conocimiento y opinión y entre el alma y el cuerpo. Mi excusa para discutir ese sistema, a pesar de la «Carta Séptima», es que es aquel del que la filosofía occidental es una nota a pie de página. Esas son las doctrinas que la gente ha estado discutiendo desde Aristóteles y Cicerón hasta Allan Bloom, las que les venían a la

cabeza al oír mentar el nombre «Platón». Por consiguiente, también yo las discutiré.

Las críticas a Platón en tiempos recientes no conciernen por la mayor parte a su política, sino a su concepción de los seres humanos y, particularmente, a su concepción de la verdad como correspondencia con una realidad trascendente. Germino está de acuerdo con Platón en que algunas de nuestras experiencias lo son de un ser trascendente. Concuerda también con él en que «el conocimiento no es en último término inducción, sino recuerdo», y en que «el verdadero filósofo es [...] alguien que cuenta una historia de una realidad que se mueve a través de él y de todos nosotros a medida que nos debatimos con el sentido de la vida, el sufrimiento, la muerte y la inmortalidad». Yo discrepo de él en todos esos puntos.

Los filósofos que piensan que Platón empezó con mal pie creen que no hay tal cosa como una realidad que se mueve a través de todos nosotros. No hay una «trascendencia» que experimentar o alcanzar. Ni hay algo «bueno y verdadero en lo profundo de nosotros». Desde nuestro punto de vista, el platonismo es una versión secular de la religión. Comparte las ventajas y las desventajas de la fe religiosa. La principal ventaja de ambos es que hacen la vida más fácil de soportar. Su principal desventaja es que hacen una distinción entre el mundo de la apariencia —el que está asociado al tiempo, al cuerpo y a los sentidos— y un Mundo Verdadero, el de la verdad eterna, el mundo inmaterial del espíritu. Muchos filósofos recientes piensan que la desventaja sobrepuja a la ventaja, y yo estoy de acuerdo.

El principal legado de Platón a nuestra civilización ha sido la ultramundanidad; la idea de que hay algo más allá del mundo del espacio y el tiempo, que los seres humanos de algún modo tenemos un pie en ese otro mundo, y que el verdadero conocimiento consiste en recordar esa realidad no espaciotemporal. Los principales críticos de Platón han sido quienes dicen que tanto la idea de Dios como la de un mundo verdadero han hecho más mal que bien. Esos críticos creen, y yo coincido, que sería mejor pensar que los seres humanos son criaturas meramente temporales, meramente mortales. Sería

mejor abandonar la idea de que los seres humanos tienen una naturaleza y un deber de alcanzar esa naturaleza. Sería mejor pensar que nos creamos a nosotros mismos a medida que avanzamos. Sería mejor abandonar la idea de que hay algo llamado la verdad cuyo conocimiento nos hará libres.

No hay ningún argumento fulminante y definitivo que demuestre que sería mejor abandonar esas ideas religiosas y platónicas, como no lo hay a favor o en contra de la existencia de Dios. Lo único que cabe hacer en este terreno es desarrollar visiones alternativas, modos alternativos de mirar a la situación en que se encuentran los seres humanos en el tiempo presente. Intentaré desarrollar la imagen anti-platónica deteniéndome en la idea de que los seres humanos son hacedores y no, como creía Platón, conocedores.

La filosofía de Platón se centra en un intento de tomar como modelo al matemático, no al poeta. El matemático puede dar argumentos concluyentes para sus creencias. El poeta no puede, y tampoco el crítico literario, la persona que quiere insistir en que una cierta obra de arte es mejor que otra. Platón pensaba que la clase de certeza que tenía el matemático era preferible al tipo de incertidumbre bajo la que tiene que operar el poeta, o el crítico, o el político. Él se impuso la tarea de preguntar: ¿cómo deberían ser los seres humanos si fueran en principio capaces de alcanzar esa certeza?, ¿cómo deberían ser si hubiera respuestas correctas a las preguntas que Sócrates iba haciendo por ahí, preguntas sobre la naturaleza de la justicia, de la virtud y del conocimiento?

Su respuesta fue que el tiempo y el espacio debían ser a la postre una ilusión, un velo tras el cual hay algo a lo que el alma humana es afín. Igual que el cristiano cree que hay un mundo más allá de la tumba, donde veremos como somos vistos[1], así también Platón creía que el mundo en derredor no era el Mundo Verdadero, no el que podía responder las preguntas que nos veíamos abocados a hacer. A menos que hubiera alguna realidad trascendente con la que pudiéramos entrar en contacto, pensaba Platón, Sócrates habría vivido en vano. A menos que ahí afuera hubiera una verdad esperando a ser descubierta —una verdad absoluta, no relativa a las

cambiantes circunstancias humanas—, no habría modo de que los seres humanos llegaran a ser libres.

La visión contrapuesta, cuyo mejor ejemplo es Dewey, es que lo que dio propósito a la vida de Sócrates no fue descubrir una verdad permanente y absoluta, sino hacer que la gente siguiera pensando, que siguiera inventando, abrir su imaginación a alternativas. El propósito del cuestionar socrático no es poner las miras en una realidad trascendente, sino sencillamente hacer que la gente pregunte: ¿son nuestras costumbres e instituciones las mejores que podemos imaginar, o podemos hacerlo mejor? En la visión de Dewey, no hay un criterio intrínseco de ser mejor. No tenemos ningún recuerdo de la naturaleza de la bondad que nos sirva para juzgar entre las alternativas que se nos proponen. Lo único con lo que podemos comparar las costumbres e instituciones nuevas es con las viejas, del modo experimental y tentativo en que comparamos amigos, empleos o entornos nuevos con los viejos. Si la experiencia nos lleva a preferir instituciones democráticas a las monárquicas, las oligárquicas o las totalitarias, no tendremos más justificación para la democracia que la de Winston Churchill. Se atribuye a Churchill haber dicho que «la democracia es la peor forma imaginable de gobierno, exceptuando todas las demás que se han probado hasta ahora».

Si a un deweyano le preguntan cuál es su criterio de verdad, lo único que puede hacer es decir: «la libre discusión, la discusión entre personas que hacen todo lo que pueden por pensar todas las alternativas posibles y luego le dan vueltas a las ventajas y desventajas de cada una». El único test de la verdad es que es la opinión que gana en un enfrentamiento libre y abierto. Pero ese test solo sirve hasta que aparece alguien con una propuesta nueva: una nueva teoría científica, un nuevo estilo artístico, una nueva institución política. Entonces habrá que reanudar la discusión. Nunca vamos a llegar a algo de lo que podamos estar seguros que es la última palabra, la verdad final, absoluta. Porque nunca podemos estar seguros de que se han sondeado todas las alternativas posibles. Nunca habrá un tiempo en el que las preguntas socráticas se hayan vuelto innecesarias.

Puedo expresar de otra manera esta visión deweyana planteando la pregunta: ¿por qué estamos tan seguros de que la verdad no puede ser derrotada en un enfrentamiento libre y abierto? La respuesta platónica es que el conocimiento es recuerdo, que reconoceremos la verdad cuando la veamos. La respuesta deweyana es que, en nuestra sociedad democrática liberal, la libre discusión se ha convertido en el criterio, el único criterio, de la verdad. Para Dewey, no hay nada más que decir sobre la naturaleza de la verdad o del conocimiento excepto que es lo que una discusión así produce. Para Platón, hay muchísimo más que decir: hay que responder a toda una gama de preguntas metafísicas sobre la naturaleza de los seres humanos y de preguntas epistemológicas sobre la naturaleza del conocimiento humano. Para Dewey, la metafísica y la epistemología son parte del legado de Platón a Occidente, un legado sin el cual nos iría mejor.

¿Cómo decidimos entre Platón y Dewey? Como dije antes, no creo que haya un argumento definitivo. Es cuestión de si uno quiere ver la humanidad como algo que está contenido en una cosa más grande que ella y a la que debe ser fiel, o como creándose continuamente a sí misma a base de crear nuevas formas de ciencia, de arte y de comunidad. La visión platónica es consoladora y la deweyana es liberadora. La visión platónica corre el peligro de alentar el dogmatismo; la deweyana, de alentar el anarquismo. Al decidir a qué filósofo seguir, estamos sopesando entre sí estas distintas ventajas y desventajas.

La razón de que en estos tiempos se le den tantos palos a Platón no es, creo, la ignorancia o el prejuicio, sino la convicción de que el paso a una cultura secular debe venir seguido ahora de un rechazo a las versiones seculares del consuelo religioso. Muchas personas piensan, como yo pienso, que en conjunto nos va mejor ahora que las polémicas teológicas han sido separadas de las políticas. Consideran que la Ilustración liberó a las preguntas sobre cómo lidiar con la injusticia y el sufrimiento de las preguntas sobre la voluntad de Dios. Ahora les gustaría liberarlas también de las preguntas sobre la naturaleza del hombre y de la verdad. Les gustaría des-filosofizar

la política, igual que el Sr. Jefferson y los demás Padres Fundadores la des-teologizaron.

Justamente porque la filosofía occidental es una serie de notas a pie de página a Platón, liberar a la política de la filosofía es liberarla de la necesidad de prestarle más atención al platonismo.

---

[1] N. del T.: alusión a la primera carta de Pablo a los Corintios (13:12): «pues ahora vemos en un espejo, en enigma; entonces, cara a cara. Ahora conozco en parte, pero entonces conoceré como soy conocido» (según la edición de Antonio Piñero de *Los libros del Nuevo Testamento*, Madrid, Trotta, 2022). También las principales traducciones de la Biblia al inglés usan aquí «conocer» y no «ver», como curiosamente hace Rorty.

CAPÍTULO 13

EL ESTADO ACTUAL DE LA FILOSOFÍA EN ESTADOS UNIDOS

Constantemente se nos dice que, en los últimos tiempos, ha habido un rápido cambio en las disciplinas académicas agrupadas bajo el término «humanidades». A veces nos dicen también que están ahora en «crisis». Esto no es verdad de la filosofía en Estados Unidos o Gran Bretaña. En esos países la filosofía se revolucionó a sí misma haciéndose «analítica» durante la década de 1950. A partir de entonces, ha seguido siendo analítica orgullosamente y sin ningún complejo. Desde aquellas fechas no ha habido en la disciplina nada ni remotamente parecido a una guerra generacional, ni tan siquiera a una crisis. Ningún cambio estructural ha tenido lugar en ese periodo, y las perspectivas de cambio a gran escala parecen escasas en un futuro previsible.

La primera versión de la filosofía analítica —el empirismo lógico de la década de 1930— declaró su independencia de la tradición filosófica y de las humanidades. Insistió en que la filosofía debía dejar de verse a sí misma como vecina de la historia y la literatura y hacer todo lo posible por volverse científica. Carnap y Reichenbach estaban convencidos de que estudiar geología preparaba mejor para el trabajo filosófico que leer a George.[1] Sospechaban que, a Gadamer, su inmersión en Platón y en los poetas lo inhabilitaba para un trabajo filosófico serio. La declaración de independencia sostenía como autoevidente que hay una distinción clara y limpia entre cuestiones de lenguaje —cómo se usan o deben usarse las palabras— y cuestiones fácticas, empíricas. Aducía a continuación que Platón —y la mayoría de los demás autores del canon filosófico occidental— habían creado innumerables pseudoproblemas, en parte debido a su ignorancia de esa verdad autoevidente.

La distinción nítida empírico-lingüístico, junto con la idea de que los conceptos se podían «analizar» y «clarificar» descubriendo condiciones necesarias y suficientes para aplicar las palabras que los significaban, duró unos veinticinco años. No sobrevivió a las críticas de Quine y del segundo Wittgenstein circa 1950. Pero la retórica pre-quineana de unos «poderosos métodos analíticos» no fue remplazada por una nueva metafilosofía. Más bien, la segunda generación de filósofos analíticos abandonó tácitamente la vieja declaración sin intentar redactar una nueva. A los filósofos que en los años sesenta estaban escribiendo sus primeros libros y artículos les unía el sentido de un saber-hacer compartido, más que el compromiso explícito con un método o un programa. Lo mismo vale para la generación que está saliendo ahora de las facultades. En ambos casos, el saber-hacer es sustancialmente el mismo.

Ese sentimiento de un saber-hacer compartido, combinado con el de superioridad moral e intelectual hacia los filósofos de otras épocas y otras tierras (no digamos ya hacia los profesores de literatura)[2], ha bastado para mantener a la filosofía analítica intacta y segura de sí misma durante medio siglo. A los filósofos analíticos no se les da demasiado bien explicar a los de afuera qué hacen y por qué, pero normalmente tienen la impresión de que eso no es más perjudicial ni más bochornoso que la parecida incapacidad que aqueja a los topólogos, los estadísticos o los sanscritistas. Saben que se les da bien lo que hacen y no les preocupa demasiado explicar por qué hace falta hacerlo. Lo normal es que consideren las críticas de los de afuera como un síntoma de feroz anti-intelectualismo. Cuando las críticas proceden de quienes se autodescriben como humanistas —personas a las que les gustaría que la filosofía se pareciera más a la de los viejos buenos tiempos—, dan por hecho que lo que las anima es el recelo y el miedo típicos de mentes flojas cuando se enfrentan a la argumentación rigurosa.

No obstante, pese a la falta de programa metafilosófico o metodológico explícito, la mayoría de los filósofos analíticos seguramente se sentirían bastante cómodos con la explicación que

da David Lewis de la filosofía como una cuestión de reconciliar intuiciones en conflicto. Dice Lewis:

> Uno ya llega a la filosofía dotado de un depósito de opiniones existentes. No es asunto de la filosofía socavar o justificar en una medida importante esas opiniones prexistentes, sino solo intentar descubrir modos de expandirlas en un sistema ordenado. El análisis de la mente que hace el metafísico es un intento de sistematizar nuestras opiniones sobre ella. Lo consigue en la medida en que (1) sea sistemático y (2) respete aquellas de nuestras opiniones prefilosóficas a las que estamos firmemente apegados.[3]

Como especialistas que son en hacer que las cosas encajen entre sí sistemáticamente, los miembros de un departamento de filosofía se apuntarán diligentemente en bloque para oír hablar a los profesores visitantes de tópicos tan diversos como la reducibilidad de la conciencia a procesos cerebrales, la existencia de Dios, la naturaleza de la referencia lingüística, los derechos humanos, y las implicaciones de la indeterminación cuántica para la responsabilidad moral. Puede que todos participen vigorosamente en la discusión subsiguiente. Porque todos se ven como personas que, siempre que se eviten los vocabularios técnicos, pueden distinguir un argumento bueno de uno malo y una afirmación plausible intuitiva de una contraintuitiva. Los filósofos analíticos comparten las aptitudes de los jueces, que también deben conocer de una asombrosa variedad de cuestiones. Dada una confusa maraña de consideraciones enfrentadas, ambos grupos de especialistas son buenos desmadejando los hilos y viendo cuáles pueden tejerse en un dibujo coherente.

Lo que hace analíticos a los filósofos analíticos es dar por supuesto que ejercitar esa habilidad en alguna sub-área en concreto de la disciplina es *todo* lo que cabe esperar que haga un profesor de filosofía. Las personas que quieren más que eso de la filosofía —especialmente quienes, a despecho de Lewis, *sí* quieren que la filosofía socave o justifique intuiciones prexistentes— tienden a dudar de la filosofía analítica. En el caso típico, creen que los filósofos deberían proponer cambios de calado en nuestro modo de

pensar el mundo y a nosotros mismos, en lugar de inventar simplemente pequeñas distinciones para resolver choques entre pequeñas intuiciones. Lo que a los profesores de filosofía les parece un deseable profesionalismo, a ellos les parece como estar uncido a una noria.

Para discutir estas dudas, necesito discutir las relaciones entre la filosofía analítica y otros tipos de filosofía, y entre ella y otras áreas de la cultura norteamericana contemporánea. Lo haré bajo cuatro epígrafes: la historia de la filosofía, la filosofía en países no anglófonos, otras disciplinas universitarias, y debates políticos contemporáneos.

I
La historia de la filosofía

La diferencia más conspicua entre la filosofía en Estados Unidos antes de la Segunda Guerra Mundial y en el presente es el papel de la historia de la filosofía en la formación de los graduados. En los viejos tiempos, se esperaba de un estudiante de grado que adquiriera cierta familiaridad con Platón, Aristóteles, San Agustín, Tomás de Aquino, Descartes, Spinoza, Leibniz, Locke, Berkeley, Hume, Kant y Hegel. Hoy en día uno puede salir adelante con conocer por encima tres o cuatro de esas figuras. El desdén de Quine hacia la historia de la filosofía es público y notorio, y su comentario de que las personas que entran en filosofía son de dos tipos—los que se interesan por la historia de la filosofía y los que se interesan por la *filosofía*— refleja la actitud de muchos filósofos analíticos.

Aunque muchos graduados de hoy hacen sus tesis sobre temas históricos —más ahora que hace una década—, esos estudiantes generalmente no tienen un relato a lo grande, de largo recorrido, sobre el curso de la filosofía desde los griegos hasta el presente. Tienden a especializarse en un periodo o escuela, igual que los profesores de literatura se especializan en un siglo. Si sabes mucho de Berkeley o de Kant, por ello mismo no se esperará que tengas nada interesante que decir sobre Platón o Hegel.

Más en general, para un estudiante de grado de filosofía es mucho más importante saber qué está pasando ahora mismo que estar familiarizado con el canon. Así que ese cambio que dicen que últimamente se está dando en los departamentos de literatura, en los de filosofía se produjo hace bastante tiempo. Las «estrellas» de entre los que se doctoran en filosofía analítica son quienes trabajan sobre los tópicos hoy candentes, los que con más intensidad se discuten en los departamentos más prestigiosos para graduados. Esos tópicos normalmente se pueden discutir sin necesidad de conocer mucho el pasado.

Este declive del prestigio de la historia de la filosofía ha venido acompañado de un consenso en que el conocimiento de idiomas no es especialmente importante en un filósofo. Existe la impresión general de que un porcentaje abrumadoramente alto de la buena filosofía se escribe en inglés, y también la expectativa general de que, con el tiempo, la filosofía de todos los países se hará analítica. Por más que los filósofos norteamericanos de la primera mitad de siglo no fueran particularmente cosmopolitas, ahora lo son sin duda mucho menos.

II
La filosofía fuera del mundo anglófono

En Francia, Alemania, Italia y España, la historia de la filosofía aún ocupa en el currículum el lugar que ocupaba en Norteamérica antes de la Segunda Guerra Mundial. En esos países, el debate sobre la significación de los autores canónicos es todavía una forma imprescindible de comunicación entre los filósofos. Habermas y Derrida, las dos figuras principales de la denominada filosofía continental, tienen relatos a lo grande, de largo recorrido, sobre el decurso de la filosofía desde los griegos. La dirección de su obra se vio determinada en gran medida por su reacción a los relatos de Hegel, Marx, Nietzsche y Heidegger.

No puedes ser muy buen lector de Derrida o Habermas si no estás familiarizado con estos últimos autores. De manera típica, los filósofos analíticos no lo están. El resultado es que, en Estados

Unidos, Habermas se enseña más en los departamentos de ciencia política, y Derrida en los de literatura, que en los de filosofía. Hay toda una tradición de pensamiento filosófico europeo, que abarca los dos últimos siglos, que raramente se menciona en los departamentos de filosofía norteamericanos.

Aunque el *establishment* analítico no trata con tanto desdén a los filósofos no analíticos como hace quince años, todavía se siguen alzando cejas cuando un graduado de uno de los grandes departamentos de filosofía norteamericanos se interesa por Heidegger o Derrida. Si ese o esa estudiante pasa a hacer su tesis sobre cualquiera de los dos autores, sus posibilidades de conseguir un buen puesto docente disminuyen. Por lo general, los departamentos de filosofía contratan a una persona que enseñe filosofía continental (debido a la demanda de los estudiantes de grado), pero rara vez a más de una. En cambio, lo probable es que tengan a tres o cuatro personas que trabajan en filosofía analítica del lenguaje y de la mente.

La desatención al pasado y a los extranjeros está hoy tan profundamente arraigada en los departamentos de filosofía norteamericanos y británicos que parece probable que la división analítico-continental se perpetúe indefinidamente. A fuerza de cambiar la idea de la gente de lo que cuenta como trabajo filosófico respetable y de lo que tiene que saber uno para ser filósofo, cabe que la filosofía analítica se haya hecho inmortal. Creo probable que «filosofía» continúe denotando durante unos pocos siglos todavía dos empresas completamente distintas y geográficamente separadas.

III
La relación de la filosofía analítica con otras disciplinas

La filosofía analítica arrancó como un intento de sacar a la filosofía de «las humanidades» y ponerla en la senda segura de una ciencia. Ese intento vino acompañado del de explicar y ensalzar la naturaleza y métodos distintivos de las ciencias naturales. En la

década de 1950, la filosofía de la ciencia estaba en el núcleo de la filosofía analítica, y el contraste entre Carnap y Heidegger se veía como el de un enfoque científico de la filosofía frente a uno acientífico y, por consiguiente, irracionalista, sesgado en contra de la democracia y a favor del fascismo.

Los propios científicos, sin embargo, no se interesaban especialmente por la filosofía de la ciencia. Los intentos de vincular a la filosofía con ciencias naturales o sociales concretas no han dado muchos frutos, salvo quizá la cooperación entre filósofos y psicólogos cognitivos en investigaciones sobre inteligencia artificial. Como los filósofos ya se habían desconectado de las humanidades, en particular de la historia y la literatura, el resultado ha sido que la filosofía se ha convertido en una disciplina de lo más aislada. La mayor parte de los profesores de universidad tienen escasa idea de lo que hacen sus colegas de filosofía y poco interés en averiguarlo.

Este aislamiento se ha visto solo ligeramente mitigado por la irrupción, en las décadas de 1970 y 1980, de programas de «filosofía aplicada», particularmente ética aplicada. Ahora hay programas bien asentados de ética médica en las facultades de medicina y de ética de los negocios en las de empresariales, en gran parte impartidos por doctores en filosofía. Pero los filósofos adscritos a ellos pierden contacto rápidamente con el departamento de filosofía de sus universidades. La literatura en esos campos de la ética aplicada es a estas alturas tan vasta que a quienes trabajan en ellos les queda poco tiempo para mantenerse al día de otras áreas de la investigación filosófica.

IV
Debates políticos contemporáneos

Sea cual sea la situación en otras disciplinas tradicionalmente «humanísticas», no ha habido una politización de la filosofía. En la década de 1960 los profesores de filosofía fueron activos en el movimiento contra la guerra de Vietnam, y la *American Philosophical Association* aprobó resoluciones contra la guerra. Pero la idea de que la filosofía tenga más relevancia en las luchas por la justicia social

nunca ha prendido realmente entre los filósofos analíticos. Su retórica sigue siendo la de una búsqueda pura, cuasi-científica, desinteresada, de la verdad.

La filosofía política es una sub-área en auge, pero en buena medida se dedica a poner en orden nuestras intuiciones, por ejemplo, sobre el peso relativo que hay que dar a la libertad y la igualdad a la hora de modelar las instituciones sociopolíticas. En los departamentos de filosofía de Estados Unidos prácticamente nadie cuestiona seriamente la tradición del liberalismo político. La mayoría de los profesores de filosofía política del país se identificarían con el ala izquierda del Partido Demócrata.

El cambio de estilo y de problemática que caracterizó a buena parte de la filosofía analítica no produjo, sin embargo, ningún cambio significativo en la filosofía política en lengua inglesa; Rawls, por ejemplo, habría escrito más o menos el mismo libro si el giro lingüístico en filosofía jamás hubiera tenido lugar. En este momento, la única diferencia entre un teórico político formado en un departamento de ciencia política y uno formado en el de filosofía es que el primero seguramente estará un poco más familiarizado con la historia de su materia.

La única iniciativa sociopolítica que ha sido importante para la filosofía anglófona en los últimos años quizá sea el feminismo. Como ocurre en muchos otros campos de estudio, algunos de los libros de filosofía más interesantes en nuestros días tienen un enfoque feminista. Pero la mayoría de ellos están escritos por mujeres cuyos filósofos preferidos son continentales, no analíticos. Autores como Derrida han indicado cómo la tradición filosófica occidental refleja supuestos específicamente masculinos y han tratado de cambiar las intuiciones a que ha dado lugar esa tradición. La filosofía analítica no ha sido particularmente útil para las feministas.

* * *

Hasta aquí las descripciones. Paso ahora a la cuestión de si, como se ha insinuado a menudo, el creciente aislamiento de los departamentos de filosofía analítica es síntoma de que esa filosofía

se ha convertido en una especie de escolástica decadente. Hilary Putnam, por ejemplo, cita el mismo pasaje de David Lewis que he reproducido antes y reacciona diciendo que «las intuiciones a las que el propio David Lewis confiere peso —por ejemplo, la de que, si las propiedades son unidades simples, entonces es ininteligible cómo una propiedad puede "implicar" otra propiedad— me parece que distan mucho de tener significación práctica ni espiritual».[4] Putnam comenta que «el método que recomienda Lewis era el método de los filósofos en la Edad Media».[5]

No hay duda de que Putnam tiene su punto de razón. Una vez que los filósofos deciden que no es asunto suyo socavar ni justificar las intuiciones existentes, no cabe esperar que el resto de la cultura les preste demasiada atención. El desespero de Putnam con la situación actual de la filosofía analítica lo comparten muchos filósofos de países no anglófonos. Normalmente esos extranjeros admiran el *esprit de corps* de los filósofos analíticos, pero no están seguros de que lo que escriben necesite leerse.

Por otro lado, siempre es difícil distinguir entre un escolasticismo indeseable y un deseable profesionalismo. Los filósofos analíticos no serán muchas cosas, pero ciertamente son profesionales competentes, trabajadores, entusiastas. Aunque, claro está, también lo eran los ockhamistas y escotistas del siglo xv. Desde el punto de vista de la libertad y la autonomía académicas, se diría que, cuando aparece un grupo así de profesionales, el resto de la universidad y de la cultura debe dar por supuesto que saben lo que están haciendo y debe dejarles en paz para que sigan con ello. Desde una reflexión histórica, en cambio, se diría que ha habido largos periodos en que una disciplina humanística ha estado simplemente marcando el paso sin moverse del sitio y no ha alcanzado ningún logro de interés.

Mi propia sensación es que intentar poner a la filosofía en la senda segura de una ciencia fue un error, pero que también lo es preguntar si determinada disciplina viene cumpliendo con el trabajo que se le ha asignado. Me inclino a decir que la filosofía analítica ha perdido fuelle en cuanto que movimiento (como se empezó a decir

de la Nueva Crítica hacia 1960). Pero la verdad es que los movimientos, y de hecho las disciplinas, tampoco importan tanto. La Nueva Crítica produjo varios cientos de miles de libros y artículos anodinos y tópicos, pero la gente seguirá leyendo durante bastante tiempo con interés y placer a Empson y Eliot, a Brooks y a Tate. La filosofía analítica ha producido casi el mismo número de escritos tópicos, predecibles, alicortos y endogámicos, pero todavía la gente podrá leer algún tiempo más con placer e interés a Ryle, Sellars, Davidson, Dennett y otros pocos.

En las humanidades, como en las artes, el buen trabajo no lo producen los métodos, o los movimientos, o los equipos, sino una imaginación idiosincrásica. Da igual qué esté de moda en los departamentos de Literatura —si el descubrimiento de la forma orgánica, o el de que todas las formas se caen a pedazos—, siempre habrá un ocasional Harold Bloom o un Kenneth Burke, una esporádica mente original. Tales mentes están tan bañadas en los libros que analizan, y en lo que esos libros han hecho para darles a ellos un sentido de quiénes son, que en su trabajo influye poco qué modas haya.

Incluso si, dentro de cincuenta años, el noventa y nueve por ciento de los problemas que discuten los filósofos analíticos del lenguaje contemporáneos parecen artificiales y verbales, todavía merecerá la pena leer unos pocos libros salidos de esas discusiones. Russell y Moore son tan aburridos y anticuados como pueda llegar a serlo un filósofo, pero si Wittgenstein no hubiera leído sus obras nunca habría escrito las *Investigaciones filosóficas*. Hoy nadie quiere leer a Christian Wolff, pero sin él no habríamos tenido la *Crítica de la razón pura*. Nadie lee a los filósofos de Oxford del 1600, pero si no hubieran sido tan decadentes como fueron, puede que nunca hubiéramos tenido las ásperas polémicas del *Novum Organum* o la sección del *Leviatán* de Hobbes llamada «Los invernaderos de la vana filosofía»[6]. Tal vez un movimiento ha cumplido ya con lo suyo si produce tres o cuatro manifiestos inflamados e interesantes, y provoca (una o dos generaciones después) tres o cuatro inflamadas y sugerentes reacciones.

[1] Hemos sido incapaces de determinar con certeza a quién se refiere Rorty aquí. Una posibilidad es el economista y periodista norteamericano Henry George. —Eds. [N. del T.: una segunda posibilidad, sugerida por el profesor italiano Roberto Gronda a los editores con posterioridad a la publicación del libro, es que se trate del poeta alemán Stefan George (1868-1933). Esta conjetura parece más lógica, pues remite a la cultura germánica a la que pertenecían Carnap y Reichenbach y, sobre todo, enlaza con la referencia a Gadamer y «los poetas» de la frase siguiente (hay un ensayo de Gadamer como contribución al simposio *Die Wirkung Stefan Georges auf die Wissenschaft*, de 1985).]

[2] Uno de mis antiguos colegas en Princeton empezó un libro con la siguiente observación: «si, como a menudo se dice, básicamente existen solo dos tipos de personas —los positivistas lógicos y los malditos profesores de literatura—, entonces yo soy un positivista lógico».

[3] David Lewis, *Counterfactuals*. Oxford: Blackwell, 1973, p. 88.

[4] Hilary Putnam, *Renewing Philosophy*. Cambridge, Massachusetts: Harvard University Press, 1992, p. 139 [hay trad. esp.: *Cómo renovar la filosofía*, Madrid, Ed. Cátedra, 1994].

[5] *Ibid*., p. 136.

[6] N. del T.: esa sección lleva el título «De las tinieblas derivadas de la vana filosofía de tradiciones fabulosas» (*Leviatán*, Cuarta Parte, capítulo 47, en la traducción de Carlos Mellizo para Alianza Ed.). No parece que Rorty lo cite mal por error; lo probable es que esté utilizando una variante humorística (suya o de su ambiente académico), pues lo que Hobbes denuncia ahí es —en parecido espíritu al de Bacon— el saber anquilosado, abstracto, libresco, «decadente» en suma, de las escuelas, lo cual bien puede describirse como una «filosofía de invernadero».

CAPÍTULO 14

EL CONVERSACIONISMO DE BRANDOM: DAVIDSON Y *HACERLO EXPLÍCITO*

El muy ambicioso, impactante, largo y complicado libro de Brandom ofrece lo que él llama una «visión unificada de lenguaje y mente». Para dominar el libro y la terminología que despliega, uno tiene que haberlo enseñado en clase un par de veces. Como yo solo lo he leído una vez, todavía no lo he enseñado y aún estoy digiriéndolo, no estoy en condiciones de dar cuenta equilibrada de su mérito. Así que me limitaré a tratar de situar tentativamente sus conclusiones en el espacio filosófico contemporáneo. Eso significa dejar de lado la ruta inferencial, extraordinariamente elegante y sorprendentemente original, por la que Brandom llega a ellas.

Por especificar más, voy a decir qué le parece *Hacerlo explícito* a alguien que comparte con su autor la convicción de que el trabajo de Wilfrid Sellars es central en los avances filosóficos de este siglo, pero que ha sido influido por Davidson no menos que por Sellars. Como sellarsiano que ha intentado extraer las implicaciones de la obra de Davidson, me veo continuamente comparando y contrastando lo que hace Brandom con lo que hace Davidson. El trabajo de ambos me parece complementario y poco dado a chocar mutuamente, aunque en énfasis y terminología son muy diferentes. Los dos se están dedicando a completar la sustitución del empirismo por el naturalismo que ya iniciaron «El empirismo y la filosofía de lo mental», por un lado, y «Dos dogmas del empirismo», por otro. En esa empresa, Brandom es a Sellars lo que Davidson a Quine.

En concreto, tanto Brandom como Davidson intentan lo que el primero describe como «explicar el decir, juzgar o saber-*que* proposicionalmente *explícitos* en términos de habilidades, capacidades o un saber-cómo prácticamente implícitos».[1] Ver la racionalidad y la cognición como una cuestión de términos de un

complejo saber-cómo está en consonancia con la propuesta de Davidson de que abandonemos la idea del lenguaje como si tuviera *estructura* y borremos «la frontera entre saber un lenguaje y saber cómo movernos en el mundo en general».[2] Cuando ese paso se combina con el de abjurar de doctrinas atomistas residuales (el infortunado intento skinneriano de Quine de interponer «significados estimulativos» entre los juicios y el mundo, por un lado, y el infortunado entusiasmo de Sellars con el figurar tractariano, por otro), lo que sale es una visión no empirista, íntegramente holista, de la mente y el lenguaje. Su principal virtud es que evita lo que Davidson llama «el tercer dogma del empirismo», la distinción entre esquema y contenido. No parte de la imagen de un sujeto que impone esquema y un objeto que produce contenido y que se cruzan a medio camino en un lugar llamado «la experiencia», el lugar donde el esquema propuesto por la espontaneidad se encuentra con el contenido suministrado por la receptividad. En vez de ello, parte de personas haciendo cosas y pregunta qué cosas concretas tienen que hacer para que se diga que usan lenguajes y tienen mentes.

Brandom termina el libro diciendo que lo que acaba de darnos —«una teoría expresiva del lenguaje, la mente y la lógica» que explica las tres cosas en términos de saber-cómo— es «una teoría sobre quiénes *somos*».[3] Un poco antes ha dicho que «la conversación es el sumo bien de las criaturas discursivas».[4] Una de las asas por donde puede agarrarse su gigantesco libro es comparar su conversacionismo con la idea tradicional de que el bien sumo de las criaturas discursivas es captar correctamente Cómo es el Mundo. En otras palabras, uno puede poner en contraste la explicación de Brandom de los seres humanos como conversadores con la platónica/aristotélica que los distingue de los brutos por su capacidad de penetrar a través de las apariencias en la realidad.

La segunda es la que dice que, aunque los seres humanos parten de descripciones de la realidad que dependen de necesidades e intereses humanos, al final pueden trascenderlas y continuar hasta algo parecido a lo que Bernard Williams denomina una «concepción

absoluta de la realidad». Podemos llamar «realismo» a eso si, como yo, se está de acuerdo con Platón, Aristóteles y David Lewis en que, como dice este último, «el realismo que reconoce como empresa no trivial la de descubrir la verdad sobre el mundo necesita del realismo tradicional que reconoce identidades y diferencias objetivas, cortes en el mundo, clasificaciones discriminativas que no sean invención nuestra».[5]

Ser un realista en este sentido es exigir que toda explicación filosófica del lenguaje, la mente y la lógica divida el discurso entre una parte en la que estamos tratando de captar algo correctamente y donde, por tanto, estamos bajo lo que Crispin Wright llama «control cognitivo», y las áreas en las que estamos *solo* conversando. En la primera, las clasificaciones discriminativas proceden del mundo, y en la segunda, de nosotros. Es el tipo de división que planteó Locke y que revivieron los positivistas lógicos. A filósofos contemporáneos como Wright y John McDowell, que dan valor a lo que denominan «nuestras intuiciones realistas», todavía les tienta semejante división del discurso. Pero cualquier división así requiere dar cuenta de una relación especial palabra-mundo, como la de «representación exacta» o la de «correspondencia», que separe, por ejemplo, la física objetiva de la estética subjetiva.

Davidson, que trata a la física y la estética como semánticamente a la par, ha sugerido que sería bueno prescindir de la idea de que el lenguaje y la mente representan algo, ya que *cualquier* uso del concepto de representación parece condenado a abocarnos, primero al escepticismo y luego al relativismo.[6] Davidson quiere abandonar el tema como tal de la supuesta independencia del mundo respecto de la mente, soslayando así los debates sobre realismo y antirrealismo. Su rechazo de la distinción entre esquema representacional y contenido representado equivale a proponer que dejemos de indagar qué clasificaciones discriminativas son del mundo y cuáles son nuestras, cuáles son parte del esquema que representa y cuáles del contenido representado, como están obligados a hacer los realistas.

También Brandom quiere librarse de los dualismos tradicionales. Dice que «las prácticas discursivas incorporan cosas efectivas» y «no deben verse como si estuvieran huecas, esperando a que las rellenen cosas; son [...] tan concretas como la práctica de clavar clavos con un martillo».[7] Insiste en que su propia distinción entre norma y hecho no es un dualismo de esquema subjetivo y contenido objetivo, sino solo un modo de enfatizar que necesitamos dos maneras de describir un único conjunto de interacciones entre los seres humanos y el resto del mundo.[8] «Los hechos», dice, «no son otra cosa que afirmaciones verdaderas», porque «no existe un ámbito estrictamente no-normativo, un ámbito al que los conceptos no se apliquen».[9]

Pero él no desdeña tanto como Davidson la idea de representación, cosa que lamento. Reconoce que el «problema de la objetividad quizá sea el reto conceptual más serio al que se enfrenta el intento de basar en la práctica social las propiedades que rigen el uso de conceptos»,[10] un intento que rompe con «primitivos representacionales» tales como la «experiencia perceptual». Así que dedica su último capítulo ha argüir que su teoría puede «explicar la dimensión representacional del pensamiento y el habla».[11] La tesis de ese capítulo final es que «la dimensión *representacional* del contenido proposicional se la confiere al pensamiento y al habla la dimensión *social* de la práctica de dar y pedir razones».[12]

No obstante, la idea de que el contenido proposicional tiene dimensión representacional se reduce en Brandom a afirmar nada más que nuestras prácticas conversacionales instituyen «normas implícitas conforme a las cuales se considera que la verdad de las afirmaciones y el correcto uso de los conceptos responde a cómo son objetivamente las cosas, no a cómo son subjetivamente o, incluso, intersubjetivamente».[13] Brandom se propone mostrarnos «cómo toma cuerpo la noción de los contenidos proposicionales como condiciones de verdad, como dependientes para su verdad de los hechos acerca de los objetos que representan».[14] Toma cuerpo en el hecho de que «compromisos que fueron asumidos sobre el trasfondo doxástico de otros compromisos, disponibles para el uso

como hipótesis auxiliares, pueden aceptarse y estar a disposición como premisas sobre un trasfondo doxástico diferente».[15] Ese modo de usarlos resulta en las adscripciones *de re*, y Brandom piensa que la capacidad de hacerlas es todo lo que se necesita para distinguir la objetividad de la intersubjetividad. Como él dice, «locuciones como "de", "acerca de" y "representa" desempeñan el papel expresivo de locuciones representacionales en virtud de su uso en especificaciones *de re* de los contenidos de adscripciones de actitudes proposicionales».[16]

Sin embargo, los filósofos con fuertes intuiciones realistas van a decir que, en la idea de representar X tal como realmente es, hay mucho más que la capacidad de decir que todos estábamos equivocados al pensar que X tenía la propiedad P. Tenderán a insistir en que el inferencialismo de Brandom ciertamente puede explicar «acerca de» y «de», pero que representar la realidad es algo más que ser acerca de ella. El propio Brandom admite que «la noción de creencia *de re* que ha ocupado a los filósofos desde Quine tiene más brío epistémico que la tenue noción que se reconstruye aquí». Eso es así, continúa diciendo, porque la primera noción «entraña estar *en rapport* con un objeto particular en un sentido más fuerte que tener meramente una forma u otra de denotarlo».[17]

Brandom no quiere tener nada que ver con la idea de que «las expresiones directamente referenciales, paradigmáticamente los deícticos, posibilitan un tipo fundamental de contacto cognitivo con los objetos del pensamiento, un tipo de creencia relacional que no está conceptualmente mediada, en la que los objetos están directamente presentes a la mente».[18] Porque él es fiel al nominalismo psicológico de Sellars, la doctrina de que todo percatarse de algo, incluso de particulares, es un asunto lingüístico (y, por ende, conceptualmente mediado). No simpatiza con la tentativa de Kripke de purgar de deicticidad el realismo a base de interpretar las adscripciones *de re* como si se acoplaran a los objetos mediante una relación de referencia que pende suelta de las inferencias hechas a o desde esas adscripciones.

La diferencia entre Brandom y Davidson, por un lado, y los sedicentes defensores de nuestras intuiciones realistas, por otro, se puede sintetizar diciendo que los primeros reducen todas las relaciones palabra-mundo *que no son la causalidad* a relaciones palabra-palabra, conversacionales, en tanto que los segundos insisten en que hay al menos una de esas relaciones —la representación, la correspondencia, o lo que Wright llama «estar bajo control cognitivo»— que no puede reducirse a relaciones ni causales ni conversacionales. Entonces, se podría decir que la única diferencia interesante entre Davidson y Brandom es que Brandom habla de normas y Davidson de regularidades conductuales.

Esta última diferencia, no obstante, parece insignificante. Considérese la afirmación de Davidson de que «la fuente última tanto de la objetividad como de la comunicación es el triángulo que, al relacionar hablante, intérprete y mundo, determina los contenidos del pensamiento y del habla».[19] Compárese con la de Brandom de que la única diferencia entre objetividad e intersubjetividad es que tenemos en las adscripciones *de re* un modo de decir que todos han estado siempre equivocados en una o más de sus creencias sobre X. Si Brandom aceptara la tesis de Davidson de que difícilmente pueden haber estado todos equivocados en *la mayoría* de sus creencias sobre X, parece que los dos tienen poco sobre lo que discrepar.

Más adelante volveré sobre la cuestión de si Brandom aceptaría o no esa tesis. Pero antes quiero incidir sobre algunos paralelismos más entre ellos. Brandom cita con aprobación la afirmación de Davidson de que «uno no puede ser un pensante si no es un intérprete del habla de otros».[20] Davidson, creo, aprobaría lo que Brandom denomina «fregeanismo táctil», la idea de que nuestra práctica nos pone en contacto con hechos y con los conceptos que los articulan.[21] Hasta donde yo veo, no tendría nada que objetar a la afirmación de Brandom de que ya se ha dicho lo suficiente sobre la objetividad una vez hemos explicado que somos capaces de entender la posibilidad de una creencia falsa perpetua y unánime.

Este acuerdo es posible porque ni Brandom ni Davidson, si les he entendido bien, le dan uso alguno a la idea de «cómo es en sí

mismo el objeto» ni a la de «correspondencia con la naturaleza intrínseca del objeto». Los dos explican la noción de objetividad sin referirse a ellas. Así que, hasta donde se me alcanza, ninguno de los dos puede o debe concederle al realista su afirmación de que hay un Modo de Ser del Mundo previo a la aparición de la mente y el lenguaje. Para ambos, la idea de distinguir entre los cortes en el mundo que coinciden con las actuales clasificaciones discriminativas lingüísticas y los que no coinciden es, a lo sumo, un gesto hacia un posible futuro en que los seres humanos desplegarán clasificaciones lingüísticas discriminativas más útiles que las que estamos haciendo ahora. Es un contraste entre prácticas discursivas más pobres y más ricas, no la distinción que hace David Lewis entre unas clasificaciones discriminativas que son del mundo y otras que son nuestras. Brandom y Davidson presumiblemente dirían que, antes del lenguaje y la mente, había tantos Modos de Ser del Mundo como vocabularios descriptivos sean capaces de formular los usuarios del lenguaje, y que ninguno de ellos está más estrechamente relacionado con el mundo que los demás. En cambio, parece que Lewis, Bernard Williams y Crispin Wright se comprometen a decir que el vocabulario de la física tiene una relación con el mundo más estrecha que el de la estética. Consideran que los fotones del mundo prehumano son parte del Modo de Ser del Mundo, pero afirman que la belleza de los crepúsculos prehumanos no es parte de ese Modo.

Me parece que Brandom suscribe la doctrina de Davidson de que la mayoría de las creencias que alguien tiene deben ser verdaderas. Se diría que esa afirmación está implícita en la doctrina del propio Brandom de que capacidad lingüística, racionalidad y capacidad *cognitiva* son tres formas de describir una única aptitud.[22] Eso parecería implicar que la idea de sistemas de creencias predominantemente falsas carece de sentido, y que, en términos de Davidson, «la creencia es verídica en su naturaleza». Para ambos, decir algo verdadero no debe verse como el resultado de una deseable relación no causal entre una descripción y un mundo no descrito —relación que solo se da si los dos términos relacionados dividen de la misma manera—, igual que ganar en un juego no debe

verse como una relación entre una práctica social y algo exterior a esa práctica. Entiendo que, para los dos, las únicas relaciones relevantes entre lo descrito y la descripción, aparte de la de ser-acerca-de, son las causales. Esas relaciones causales son, con mucho, demasiado complejas para describirse en términos de coincidencia, correspondencia o encaje. Adscribir verdad, como adscribir creencia, entraña un holismo que resiste cualquier intento de cortar lo que no es lenguaje en trozos a-tamaño-de-hechos que estén en relaciones causales con oraciones.

No obstante, puede parecer que el enfoque tarskiano de Davidson sobre la semántica se aparta mucho de la insistencia de Brandom en que «las locuciones representacionales [como "verdadero" y "refiere"] no sirven para desempeñar el papel de elementos primitivos de una teoría semántica».[23] Es fácil imaginar a Davidson retrocediendo ante la afirmación de este de que «la verdad y la referencia son ficciones de los filósofos generadas por malentendidos gramaticales.[24] Él dice explícitamente que el predicado de verdad de Tarski es un predicado «legítimo»,[25] mientras que Brandom sostiene expresamente que solo negando que sea un predicado se pueden esquivar las paradojas del deflacionismo.[26]

Sin embargo, no estoy seguro de que estas diferencias marquen diferencia alguna. Porque ambos concuerdan en que la palabra «verdadero» es indefinible. Los dos discreparían de críticos de Tarski como Hartry Field, y por las mismas razones. Los dos piensan que nunca vamos a entender mejor de lo que ya entendemos cómo enganchan las palabras con el mundo, y que no necesitamos nada parecido a una teoría nueva, mejorada, de cómo opera la verdad o la referencia. Los dos ridiculizan la idea de unos «hacedores-de-verdad», unos trozos de mundo que se arriman a las oraciones haciéndolas así merecedoras de ser creídas.[27] Creo que Davidson podría concederle encantado a Brandom que no existen cosas tales como hechos semánticos, porque «dominar un vocabulario semántico solo nos da una forma nueva [...] de llegar a un campo de hechos no semánticos al que ya teníamos acceso».[28]

Pero, cabría protestar, seguro que Davidson jamás podría estar de acuerdo con la afirmación de Brandom de que «uno no debe explicar el poseer contenido proposicional en términos de condiciones de verdad».[29] Nótese, no obstante, que Brandom no está negando que cualquier cosa que tenga un contenido proposicional tiene condiciones de verdad.[30] Solo dice que lo que le da ese contenido y esas condiciones es la misma práctica social. Respecto de este asunto, la claridad filosófica consiste en decir que tanto la lógica como la semántica no son sino modos de llamar la atención hacia rasgos de esa práctica. Nótese también que Davidson no está por responder a la pregunta «¿por qué crees que ese ruido tiene contenido proposicional?» diciendo «porque tiene condiciones de verdad». Es de suponer que coincidiría con Brandom en que esa sería una explicación del tipo «poder dormitivo». Coincidiría con seguridad en que «esa oración es verdadera porque se corresponde con la realidad» es de ese tipo.[31]

Brandom suscribe la opinión de Davidson de que «la *única* constricción a la hora de asignar denotaciones a expresiones sub-oracionales [es] que la asignación haga aflorar correctamente las condiciones de verdad»[32], condiciones de verdad que han sido determinadas tomando nota de lo que resulta socialmente aceptable. Pensemos en una lingüista de campo que está elaborando una teoría de la verdad para un lenguaje natural, que intenta adquirir el necesario saber-cómo para pelearse con los nativos como una hermana más. Busca evidencia de normas sociales que regulen qué signos y ruidos resulta aceptable hacer en qué entornos, y a qué acciones, lingüísticas y de otro tipo, da lugar típicamente la producción de tales ruidos y signos. A medida que construye su saber-cómo interpretativo, puede convertirlo en saber-que escribiendo oraciones-T. Pero, si nunca ha leído a Tarski o a Davidson, puede hacer la misma tarea componiendo un anticuado manual de traducción tipo Berlitz. La diferencia es meramente entre usar «es verdadero si y solo si» como expresión para conectar oraciones nativas con oraciones de su lengua materna y usar «significa». No haría mejor ni peor si se olvidara por completo tanto

de la verdad como del significado y se limitara a decir «encuentran apropiado decir tal-y-tal únicamente en las condiciones en que nosotros hallaríamos apropiado decir cual-y-cual». Una vez más, la diferencia entre convertir el saber-cómo en saber-que utilizando «verdadero», utilizando «significa» y utilizando «aprobado socialmente», parece no marcar diferencia alguna.

Si a Davidson le preguntaran si él, como Brandom, considera que las locuciones veritativas [*truth-locutions*] indican fuerza, no contenido,[33] me lo imagino resistiéndose a tener que elegir. Pero yo entiendo que su opinión de que la palabra «verdadero» es indefinible e inanalizable equivale a coincidir con Brandom. Entiendo que comparte el entusiasmo de este con la insistencia de Frege en que «la verdad es indefinible, algo cuya comprensión va siempre implícita ya en el afirmar».[34] Y, hasta donde yo veo, la línea entre palabras indefinibles y definibles viene a ser la línea entre palabras que se usan para indicar fuerza y palabras que se usan para especificar contenido, entre palabras de las que basta con que elijas hacer uso y palabras cuyo uso puede explicarse de un modo no obviamente circular mediante otras palabras. Brandom piensa que, tanto términos lógicos como «y» como términos semánticos como «verdadero» indican fuerza. Por eso dice que la mejor forma de concebir «la condición de verdad que pesa sobre el conocimiento» no es afirmando que lo conocido debe tener una propiedad llamada «verdad», sino simplemente entendiéndolo como la tesis trivial de que suscribir la afirmación que hace alguien de que conoce algo es automáticamente suscribir lo que ese alguien dice que sabe.[35]

* * *

Hasta aquí mi propuesta de que a Davidson y Brandom hay que verlos como parte de los intentos de naturalizar la semántica y la filosofía de la mente. Ese proyecto culmina, a mi juicio, en la afirmación de Davidson de que, una vez admitimos que no vamos a ser capaces de definir la verdad, ya no necesitamos, y de hecho no podemos, ser realistas ni antirrealistas. Davidson lo dice así: «el realismo, al insistir en una correspondencia radicalmente no

epistémica, pide de la verdad más de lo que podemos entender; el antirrealismo, al limitar la verdad a lo que puede ser establecido, la priva de su papel como estándar intersubjetivo».[36]

Brandom coincide con Davidson en que no debemos tratar de reducir la objetividad a intersubjetividad definiendo la primera, con Peirce y Putnam, como lo que tendremos al final de la investigación. Pero sospecho que los realistas van a insistir en que, en su postura y en la de Davidson, hay implícita una reducción de ese tipo. Yo creo que hay un sentido en el cual tienen razón. Más en general, pienso que, una vez que naturalizas la semántica y la filosofía de la mente, puedes darle un significado respetable a la célebre afirmación de William James de que la verdad es utilidad para los seres humanos, y también a la de que la verdad objetiva es una cuestión de sobre qué pueden ponerse de acuerdo los seres humanos. Creo que Davidson y Brandom deberían estar más dispuestos de lo que están a admitir que James estaba en lo cierto.

El argumento estándar para equiparar la verdad a la utilidad es que palabras y conceptos son herramientas desarrolladas para servir a propósitos humanos. Así, si Brandom tiene razón en que «no hay ámbito al que los conceptos no se apliquen», entonces James tiene que tenerla en que «el surco de la serpiente humana está por todas partes», y Williams y Lewis no la pueden tener cuando insinúan que podemos hallar un ámbito libre de los surcos de esa serpiente. Debemos decir que no hay un Modo de Ser del Mundo a no ser uno de los diversos modos en que los seres humanos han encontrado útil describir el mundo, y que todos esos diversos modos están igualmente atados a necesidades e intereses humanos.

Me parece que Brandom debería conceder sin más este punto. Sin embargo, si lo hiciera tendría que revisar al menos un pasaje de su libro. Es ese en el que dice que «los hechos no lingüísticos podrían ser en gran medida los que son incluso si nuestras prácticas discursivas fueran completamente diferentes [o no las hubiera en absoluto], pues qué afirmaciones son verdaderas no depende de que las afirme nadie. Pero nuestras prácticas discursivas no podrían ser las que son si los hechos no lingüísticos fueran diferentes».[37]

En el pasaje que acabo de citar, Brandom recae en algo que se parece más que demasiado a la distinción esquema-contenido. Lo que debería haber dicho es que no hay hechos no lingüísticos, que es lo que me parece que viene a querer decir su afirmación previa de que «no hay ámbito al que los conceptos no se apliquen». Entonces podría haber dicho que es absurdo intentar distinguir nuestra contribución a los hechos y la contribución que hace el mundo, que no hay modo de separar qué parte de los hechos es esquema y qué parte es contenido. Podría hacerle justicia tanto a James como a «nuestras intuiciones realistas» coincidiendo con James en que resulta igual de absurdo decir que una creencia es verdadera porque es útil o que es útil porque es verdadera.[38]

En la misma página en que expresa la idea que acabo de criticar, Brandom dice que:

> No depende de nosotros qué afirmaciones son verdaderas (es decir, cuáles son los hechos). En un sentido, depende de nosotros qué ruidos y signos expresan qué afirmaciones, y por consiguiente, en un sentido más atenuado, cuáles expresan afirmaciones verdaderas. Pero la constricción empírica y práctica sobre nuestro arbitrario capricho es un rasgo ubicuo de nuestra práctica discursiva.[39]

Yo replicaría que la necesidad de formas cada vez más ricas y complejas de felicidad humana es un rasgo igualmente ubicuo de nuestra práctica discursiva. La física corpuscularista, la cuántica, el arte clásico o el romántico, no nos vienen del capricho arbitrario ni de la constricción empírica y práctica. Nos vienen de la interacción entre nuestros antepasados y el resto del universo. Esa interacción puede describirse igualmente bien como una búsqueda de normas sociales cada vez más amplias y mejores como medio para formas cada vez más amplias y mejores de felicidad humana, o como los desesperados intentos de una especie por adaptarse a un entorno hostil. Como ha argüido recientemente Bruno Latour, no tiene sentido intentar localizar una quiebra entre naturaleza y cultura.

El debate realismo-antirrealismo me parece fútil porque es una competición entre esas dos descripciones, entre las cuales no hay

necesidad de elegir. Brandom y Davidson se elevan por encima de este debate obsoleto casi todo el tiempo. Pero a veces, como en el pasaje que acabo de citar de Brandom y en el último párrafo de «De la idea misma de un esquema conceptual» de Davidson, descienden por un momento a su nivel. Lo hacen al ceder a la tentación de reafirmar «nuestras intuiciones realistas», distanciándose así del tipo de filósofo que, según los realistas, es una criatura de arbitrarios caprichos, fuera de contacto con la realidad.

Creo que Brandom podría evitar la tentación diciendo que, aunque al nivel de los enunciados sobre objetos particulares ciertamente hay una distinción clara y limpia entre objetividad e intersubjetividad, a medida que el discurso se hace más general se desvanece gradualmente. Una vez que empezamos a hablar de tipos naturales y de visiones del mundo globales, la distinción entre objetividad e intersubjetividad se vuelve absurda. En el caso de los tipos naturales, esto equivaldría a tomarse en serio la tesis de Davidson de que *la mayoría* de nuestras creencias sobre lo que son los árboles, las rocas, las brujas, los unicornios y los átomos deben ser verdaderas, aun cuando un número razonable de ellas puedan resultar ser falsas. En el caso del mundo como un todo, no podemos estar profundamente equivocados sobre cómo es. En particular, no podemos estar errados en lo que Kant creyó que habían errado sus predecesores: no puede haber un mundo-en-sí extremadamente diferente del mundo espaciotemporal en el que pensamos que vivimos.

Que el mundo está esparcido en espacio y tiempo, o que el oro es un metal, o que el arte romántico es un gran avance respecto del arte clásico, no lo determinan ni la subjetividad humana ni el carácter intrínseco de la realidad. Uno podría decir, al modo de Kant, que eso es porque por separado son inútiles y deben cooperar. Pero sería mejor decir que ni una ni otro existen. Deberíamos dejar que esos dos mitos que se sustentan mutuamente —el del capricho subjetivo arbitrario y el de la presión inmediata del Modo de Ser del Mundo— se aniquilen el uno al otro. Pero será imposible librarse de esta o cualquier otra forma de la distinción esquema-contenido en tanto no abandonemos la idea misma de hecho no lingüístico. Hasta

que lo hagamos, la afirmación hegeliana de Brandom de que la conversación es el mayor bien de las criaturas discursivas, con la que coincido de todo corazón, parecerá superficial, doxófila y frívola.

---

[1] Robert R. Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Massachusetts: Harvard University Press, 1994, p. 135 [hay trad. esp.: *Hacerlo explícito: razonamiento, representación y compromiso discursivo*, Barcelona, Ed. Herder, 2009].

[2] Donald Davidson, «A Nice Derangement of Epitaphs», en *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. de Ernest LePore. Oxford: Blackwell, 1986, pp. 445-446.

[3] Brandom, *Making It Explicit*, p. 650.

[4] *Ibid.*, p. 644.

[5] David Lewis, «Putnam's Paradox», *Australasian Journal of Philosophy* 62, n.º 3 (1984), p. 228.

[6] Véase Davidson, «The Myth of the Subjective», en *Relativism: Interpretation and Confrontation*, ed. por Michael Krausz. South Bend, Indiana: Notre Dame University Press, 1989, pp. 165-166 [hay trad. esp.: «El mito de lo subjetivo», en Donald Davidson, *Mente, mundo y acción*, Barcelona, Ed. Paidós/I.C.E.-U.A.B., 1992, pp. 51-71].

[7] Brandom, *Making It Explicit*, p. 332.

[8] *Ibid.*, p. 624.

[9] *Ibid.*, pp. 622 y 624.

[10] *Ibid.*, p. 137.

[11] *Ibid.*, p. 495.

[12] *Ibid.*, p. 496.

[13] *Ibid.*, p. 498.

[14] *Ibid.*, p. 517.

[15] *Ibid*.

[16] *Ibid.*, p. 584.

[17] *Ibid.*, p. 548.

[18] *Ibid.*, p. 551.

[19] Davidson, «The Structure and Content of Truth», *Journal of Philosophy* 87, n.º 6 (1990), p. 325.

[20] Brandom, *Making It Explicit*, p. 629.

[21] *Ibid*., p. 632.

[22] *Ibid*., p. 203.

[23] *Ibid*., p. 285.

[24] *Ibid*., p. 324.

[25] Davidson, «The Structure and Content of Truth», p. 285.

[26] Véase Brandom, *Making It Explicit*, pp. 326-327.

[27] Véase *ibid*., p. 328.

[28] *Ibid*., p. 329.

[29] *Ibid*.

[30] *Ibid*., más abajo.

[31] Véase *ibid*., p. 330; compárese con Davidson, «The Structure and Content of Truth», p. 305.

[32] Brandom, *Making It Explicit*, p. 364.

[33] Véase *ibid*., p. 296.

[34] *Ibid*., p. 112.

[35] Véase *ibid*., p. 297.

[36] Davidson, «The Structure and Content of Truth», p. 309.

[37] Brandom, *Making It Explicit*, p. 331; pasaje que no está en consonancia con la p. 624, ni siquiera con las pp. 332-333.

[38] Véase William James, *Pragmatism*, p. 204. [N. del T.: en el original no se consigna a qué edición de *Pragmatism* corresponde el número de página de esta referencia; en la edición española de la obra (*Pragmatismo*. Madrid, Alianza Ed., 2016), la afirmación de James a que alude Rorty aparece en la p. 206.]

[39] Brandom, *Making It Explicit*, p. 331.

CAPÍTULO 15

EL HILEMORFISMO DE McDOWELL Y EL NATURALISMO CRUDO

He aquí una visión crudamente naturalista de la relación entre mente y mundo: una mente es un entramado muy grande de estados intencionales atribuidos a un objeto de muy compleja conducta, a una cosa que produce outputs que es provechoso traducir a oraciones de nuestro lenguaje. Es provechoso si resulta que la atribución de estados intencionales que la traducción posibilita es de mucha ayuda, casi algo imprescindible, para predecir la conducta del objeto en cuestión.

Esos estados intencionales son un elemento en las interacciones causales del objeto (una máquina, por ejemplo, o un organismo) con su entorno. Hallamos que podemos predecir y controlar más fácilmente tales interacciones describiendo el objeto como teniendo creencias y deseos —y, por consiguiente, usando en la descripción las nociones de «ser-acerca-de» y «verdad» además de la de «causa»— que atribuyéndole solo estados no mentales, digamos, de órganos o de circuitería. Pero que esas dos nociones adicionales sean útiles no proporciona de suyo una razón para emplear las de «representación», «intuición», «concepto», «receptividad» o «espontaneidad».

Por ejemplo, cuando interpretamos que los signos y ruidos que produce un cerebro en una cubeta son acerca de estados del entorno con el que está interactuando (una cubeta, cables, un programa de ordenador), no hace falta que nos preocupemos de si *en realidad* está teniendo creencias acerca de un entorno imaginario (una soleada playa, por ejemplo). Ni hace falta tampoco que nosotros mismos nos preocupemos de que podamos malinterpretar completamente nuestro propio entorno o nuestros objetos intencionales. El naturalismo crudo no deja sitio a semejante

escepticismo. Porque los naturalistas crudos no preguntan si, o cómo, una mente «incide en» su entorno, o si o cómo lo «representa». Relacionan simplemente la conducta de un objeto con la nuestra de cualquier modo que nos haga lo más fácil posible predecir qué ocurrirá a continuación.[1]

He aquí una visión menos crudamente naturalista de la relación entre mente y mundo. La mente es un dispositivo para captar correctamente el mundo. Esto solo puede hacerse si el modo en que el mundo es se reproduce de alguna manera en la mente. (A modo de identidad, como en Aristóteles y Hegel, o a modo de representación, como en Descartes y Locke.) Tal reproducción solo será posible si la relación entre la mente y el mundo es más complicada de lo que la imagen crudamente naturalista considera que es. Las operaciones de la mente deben ser tales que las relaciones entre los elementos de sus estados intencionales a menudo «se correspondan con» o «representen» las relaciones entre las cosas de su entorno. Así que tenemos que describir una relación entre la parte de la mente que hace outputs en forma de expresiones (espontaneidad) y la parte que recibe inputs del mundo (receptividad) que asegure que la interacción entre ambas normalmente conserve la deseada correspondencia.

La primera visión, crudamente naturalista (sugerida, si es que no implicada, por las de Donald Davidson y Daniel Dennett), atrae a muchos filósofos contemporáneos. La segunda la expuso por primera vez Aristóteles en su explicación hilemorfista de la relación entre mente y mundo. Sin embargo, sobre qué relación hay entre las propiedades inteligibles de las sustancias y sus propiedades sensibles, y en torno a cómo la mente cuela la nata inteligible de la leche sensible, Aristóteles fue vago.

La sombra de esas preguntas se alargó aún más cuando resultó que Demócrito estaba en lo cierto sobre nuestro entorno. Entre 1600 y 1900, las explicaciones de la naturaleza abandonaron poco a poco la forma y la materia en pro de los átomos y el vacío. Aunque Aristóteles creía que estaba dando una explicación del conocimiento más naturalista que la de Platón, la naturaleza resultó parecerse

muy poco a lo que él pensaba.[2] Así que su hilemorfismo empezó a parecer estrafalario. Únicamente sobrevivió en las discusiones filosóficas sobre el conocimiento y la mente gracias a la distinción de Kant entre la forma y la materia del pensamiento. Aunque la explicación del propio Kant de los objetos espaciotemporales como intuiciones moldeadas por conceptos no produjo sino una aparatosa forma de escepticismo —la doctrina de que el mundo espaciotemporal era «meramente fenoménico»—, el posterior rechazo del idealismo trascendental no dio lugar a un repudio de la terminología hilemorfista de Kant. Los filósofos, ay, han seguido hablando de dos tipos de representaciones (conceptos e intuiciones) y de dos partes de la mente (la espontaneidad y la receptividad).

Esa forma de hablar es casi inevitable, sin duda, si se piensa en la creencia verdadera como captar correctamente algo. La imagen que, para bien o para mal, ha seguido teniendo cautivos a los filósofos insiste en que la filosofía debe responder a la pregunta de Kant sobre cómo logra la mente captar correctamente el mundo. El principal motivo para resistirse al naturalismo crudo es la convicción de que esa pregunta no se puede abandonar por las buenas, y la consiguiente inferencia de que hay que preservar algo parecido a lo que Davidson llama «la distinción esquema-contenido». Se considera que Davidson se deja fuera algo importante cuando, en «De la idea misma de un esquema conceptual», dice que «la idea de encajar con la totalidad de la experiencia, como la de encajar con los hechos, o la de ser verdadero de los hechos, no añade nada inteligible al simple concepto de ser verdadero».[3] Quienes discrepan de Davidson en este punto insisten en que las relaciones causales (incluso si se complementan con el empleo de los conceptos davidsonianos, no analizados, de ser-acerca-de y verdad) son insuficientes para describir la relación entre la mente y el mundo. Creen que necesitamos describir relaciones adicionales, específicamente *cognitivas*, relaciones que nos darán, si no un análisis de las nociones de «acerca de» y «verdadero», sí por lo menos algo interesante que decir sobre ellas.

Mente y mundo, de John McDowell, es muy útil para enfocar más claramente este desacuerdo. Porque McDowell escribe en respuesta directa y explícita a Davidson, arguye enérgicamente la insuficiencia de su punto de vista y utiliza sin rubor terminología kantiana, hilemorfista. Piensa que el naturalismo en filosofía ha ido ya demasiado lejos y espera remplazarlo por lo que él llama «un platonismo naturalizado».

McDowell le atribuye a Davidson una «plácida confianza» en que «el contenido empírico puede estar inteligiblemente dentro del cuadro incluso aunque estipulemos cuidadosamente que los impactos del mundo sobre nuestros sentidos nada tienen que ver con la justificación».[4] Piensa que, cuando Davidson argumenta que la mayoría de nuestras creencias deben ser verdaderas, simplemente «se está sirviendo de la idea de [...] una colección de estados que tienen contenido»[5] sin hacer nada por explicar cómo *pueden* tenerlo. Arguye que las teorías coherentistas de la verdad y el conocimiento del tipo que recomienda Davidson «expresan precisamente la perturbadora idea de que la espontaneidad del pensar conceptual no es susceptible de control desde afuera».[6]

Como se ve por estas citas, las críticas de McDowell a Davidson se formulan en términos de unas nociones —«contenido empírico», «control racional», «espontaneidad»— que a los naturalistas crudos les sirven de poco. Davidson, por ejemplo, quedaría perplejo al ver que se le endosa la idea de que «la experiencia no puede ser nada más que un impacto extra-conceptual sobre la sensibilidad», y la inferencia consiguiente de que la experiencia debe estar «fuera del espacio de las razones».[7] Porque Davidson no habla mucho de experiencia ni de sensibilidad: dada su visión del asunto, no le hace falta. Puede hablar simplemente de causas y efectos de las creencias y los deseos.

Puesto que yo interpreto a Davidson como lo que McDowell llama un «naturalista crudo»,[8] se me hace raro pensar que sostiene tesis sobre «nuestros sentidos» y «la espontaneidad del pensar conceptual», y me desconcierta oír que no ha tenido en cuenta la posibilidad de que «las capacidades conceptuales, capacidades que

pertenecen a la espontaneidad, estén ya operando en las experiencias mismas, no solo en los juicios basados en ellas».[9] Esta idea de unos «juicios basados en la experiencia» no es fácil de encajar en la descripción de Davidson de cómo operan las mentes, así que cuesta imaginarle concibiendo siquiera esa posibilidad que se dice que él pasa por alto.

Más en general, cuesta ver cómo a alguien que no encuentra necesario o útil el vocabulario hilemorfista de Kant se lo podría convencer de que, como cree McDowell, la clave para comprender la relación entre mente y mundo es darse cuenta de que «el contenido conceptual ya lo llevan consigo las impresiones que nos deja en los sentidos la realidad independiente».[10] Decirle *eso* a Davidson, el enemigo declarado de toda distinción esquema-contenido, parece una petición de principio para todas las preguntas de interés aquí. Usar jerga kantiana para leerle la cartilla a Davidson me suena tan chocante como decirle a Galileo que no se habría hecho tan crudamente corpuscularista si hubiera captado el punto de Aristóteles de que el dominio de la forma sobre la materia se extiende al movimiento local no menos que al cambio sustancial, y que, por consiguiente, el movimiento puede ser no menos natural que violento.[11]

Para juzgar las críticas de McDowell a Davidson, hay que averiguar si necesitamos el círculo de términos inter-definibles cuya utilidad él da por sentada. Además de «contenido empírico» y «control racional», ahí se incluyen «fricción con algo exterior al pensamiento»,[12] «incidencia del pensamiento en el mundo»,[13] «interacción racional entre receptividad y espontaneidad»,[14] «responsabilidad racional hacia el mundo»,[15] «relación de pertinencia» (entre experiencias y creencias),[16] «apariencias que constituyen razones para juicios»[17] y «libertad kantiana».[18] Para McDowell, la inutilidad de cualquiera de esas expresiones significa la inutilidad de todas. Sospecha (creo que con razón) que Davidson cuestionaría la utilidad de cada una de ellas.

McDowell piensa que «¿cómo puede tener algo contenido empírico?» es una buena pregunta. Cree que «¿cómo puede ejercer

el mundo control racional sobre la mente?» lo es también. A la primera pregunta, Davidson podría haber respondido, en un esfuerzo por encontrarse con McDowell a mitad de camino, que cualquier creencia adscrita sobre la base de una interpretación radical que comienza por observar la interacción causal de la máquina u organismo con su entorno tiene tanto «contenido empírico» como uno puede razonablemente pedir. A la segunda respondería seguramente que, si el mundo ejerce control causal al modo que nos es familiar, haciendo que la mente altere sus creencias y deseos en respuesta más o menos predecible a cambios en su entorno, eso *es* control racional. Pero tales respuestas equivaldrían en realidad a rechazar las preguntas que se están haciendo, a rechazar la exigencia de más luz que se aloja en el uso que hace McDowell de los diversos términos inter-definibles que he citado.

Quine se deshizo de un montón de malas preguntas —y, como dice Davidson, «salvó la filosofía del lenguaje como tema serio»— segregando ciertos términos inter-definibles («analítico», «sinónimo», «regla semántica», «conceptual», «cuestión de lenguaje»), términos que, en los textos de Carnap, se apoyaban los unos en los otros, y negándose a usar ninguno de ellos. Mi propia reacción hacia McDowell es que solo podemos salvar la discusión de la relación entre mente y mundo como tema serio tirando a la basura los términos inter-definibles que él utiliza para argüir la inadecuación del naturalismo crudo.

McDowell critica con cierto detalle mi gusto por tirar a la basura terminologías y preguntas. Concede que eso a veces es buena idea, como cuando dice que es apropiado encogerse de hombros ante preguntas como «¿qué constituye la estructura del espacio de las razones?», que preguntas así «no deberían darse por buenas sin mayores consideraciones, solo porque es norma que la filosofía, tal como hemos sido educados en ella, se las haga».[19] Pero piensa que yo llevo demasiado lejos el encogimiento de hombros. Afirma que me he privado a mí mismo del derecho a decir que «los problemas filosóficos sobre cómo está el pensamiento en contacto con el mundo son ilusorios», porque mi propia manera de pensar «vuelve

urgentes esos problemas» al «separar el relacionarse-con-el-mundo del entorno normativo que se requiere para dar sentido a la idea de incidencia —incidencia racional— en lo que sea». Así que mi negativa a plantearme los problemas filosóficos sobre cómo está el pensamiento en contacto con el mundo «solo puede ser un acto de voluntad, un taparse deliberadamente los oídos».[20]

Lo que a un filósofo le parece una bien justificada negativa a tomar en serio determinadas preguntas que son norma en la filosofía en que, ay, hemos sido educados, a menudo le parece a otro un deliberado taparse los oídos. Para mí Davidson nos ha dicho todo lo que necesitamos oír sobre la relación entre la normatividad —las regularidades que se exhiben en el uso del lenguaje— y el relacionarse-con-el-mundo. Lo ha hecho describiendo cómo inferimos hasta esas regularidades observando «qué episodios y situaciones del mundo causan que un agente prefiera que esta oración y no aquella sea la verdadera».[21] McDowell no cree que eso sea suficiente ni de lejos, porque nos deja sin capacidad de darle sentido a una idea —«incidencia racional»— que yo no veo que necesite tenerlo. Un filósofo que quiere eliminar un determinado círculo de términos inter-definibles (por ejemplo, los términos carnapianos que critica Quine, los términos aristotélicos que criticaba Galileo) a menudo no verá ningún problema allí donde los que consideran imprescindibles esos términos verán toda clase de ellos. McDowell, en una palabra, me parece a mí tan voluntarista como yo a él.

Entre McDowell y yo mismo (y entre él y Davidson, si entiendo bien a este último) lo que hay es una discusión frontalmente práctica sobre el éxito de nuestras respectivas terapias. Él cree que mi enfoque «hace que siga pareciendo que las preguntas filosóficas [tradicionales] *deben* ser buenas», de modo que «el resultado es que el malestar filosófico persiste, no un exorcismo de la filosofía».[22] McDowell quiere hacer exorcismo, igual que yo, y parece dispuesto a aceptar que el test de su éxito es la ausencia de todo malestar residual. Pero afirma que «el exorcismo requiere un tipo de jugada diferente que a Rorty se le da mucho peor»,[23] la

clase de jugada que él hace. Así que la única forma de decidirse entre el naturalismo crudo y el «platonismo naturalizado» de McDowell será observar la cantidad de malestar que producen nuestras respectivas prácticas exorcizadoras y terapéuticas. Ambos predecimos que el remedio del otro será peor que la enfermedad.

Concretamente, yo predigo que su versión bastante espesa del platonismo probablemente cause más malestar que mi naturalismo crudo. Él piensa que «nuestras angustias filosóficas se deben al comprensible influjo […] de un moderno naturalismo, y podemos trabajar en relajar ese influjo» platonizando el naturalismo.[24] Yo quiero aumentar el influjo hasta que produzca en el paciente justamente el resultado que McDowell cree que hay que evitar, a saber: «pensar que hay [un] problema con la idea de cómo son en cualquier caso las cosas, independientemente de lo que se certifique comunitariamente».[25] Yo creo que la idea de «cómo son en cualquier caso las cosas» requiere amputación, no tratamiento rehabilitador.

Igual que Putnam (y Davidson, creo), quiero hacer que la noción de Bernard Williams de una «concepción absoluta de la realidad» —un modo en que las cosas son en cualquier caso— parezca tan problemática como me sea posible. Trato de minimizar el malestar que causa repudiar esa noción intuitivamente plausible desgranando las consecuencias contraintuitivas de conservarla. Empiezo por interpretar «en cualquier caso» como «al margen de cualquier descripción», interpretación que la explicación de Williams de «absoluta» como «al margen de las necesidades e intereses humanos» parece autorizar. (Las descripciones son, sin duda alguna, producto de necesidades e intereses humanos.) Razono, entonces, que la idea de cómo son en cualquier caso las cosas es la idea problemática e inútil de un mundo no sujeto a descripción alguna.

Análogamente, mi estrategia para que el platonismo naturalizado de McDowell haga sentir malestar es preguntarle cómo puede él ver el mundo «como algo cuyos elementos son cosas que son el caso»[26] sin poblarlo con eso que Strawson, criticando la noción de verdad como «correspondencia con los hechos», llamó «objetos no-

lingüísticos con forma oracional». A mi juicio, la forma en que McDowell hace que el mundo ejerza sobre nosotros control racional, y no crudamente causal, consiste en darle al mundo un cuasi-lenguaje propio, un conjunto de descripciones en las que el mundo formula las cláusulas-«que» que imprime sobre nuestra sensibilidad. La afirmación de McDowell de que «el contenido conceptual ya lo llevan consigo las impresiones que nos deja en los sentidos la realidad independiente»[27] parece significar, en sustancia, que la Naturaleza está intentando decirnos algo, y que su vehículo de comunicación es nuestra facultad de receptividad.

Hasta donde yo veo, si McDowell no hubiera ya platonizado y personificado a la Naturaleza dotándola de semejante cuasi-lenguaje, de nada le valdría insistir en que nuestra «sensibilidad perceptual», al contrario que la de los mudos animales y los termostatos, está «incorporada a la esfera de la facultad de espontaneidad».[28] Por muy conceptualizada que esté mi sensibilidad perceptual, nada va a ejercer constricción racional sobre ella a menos que la fuerza que constriñe esté conceptualizada también. Entiendo que esa es la razón de que McDowell quiera «integrar el mundo dentro de lo que en Frege aparece como el reino del sentido».[29] Pues solo así puede el abismo girar la cabeza, mostrarnos un rostro cuasi-humano, imponernos demandas racionales, en vez de intimidarnos simplemente.

Coincido con McDowell en que no hay «salto ontológico entre el tipo de cosa que uno puede significar o, en general, pensar, y el tipo de cosa que puede ser el caso».[30] De hecho, considero tautológica la afirmación. Pero no creo que eliminar ese salto nos exija ver el mundo como lleno de cláusulas-«que» no lingüísticas, o no formuladas en ningún idioma humano específico, o susceptibles por siempre de formularse en cualquier idioma humano. Me parece que es esta tercera alternativa la que se deduce de su afirmación, no solo de que «*que las cosas son así y así* es el contenido de la experiencia, y puede ser también el contenido de un juicio», sino de que «*que las cosas son así y así* es también, si uno no va errado, un aspecto de la disposición del mundo; es cómo son las cosas».[31]

Para hacer que la jugada platonizante de McDowell se vea lo más espesa posible, junto las adivinanzas retóricas habituales, burlonas y anti-platónicas, cuyos ancestros se remontan a la autocrítica de Platón en el *Parménides*: cuando una facultad de sensibilidad, cara a cara con una cámara de niebla, es constreñida racionalmente al quedarle impreso *que las cosas son así y así*, ¿lo que se le imprime es que están ocurriendo un montón de colisiones de partículas elementales, o que hay un montón de fugaces líneas de un gris más oscuro sobre un fondo de un gris más claro? Cuando da contra la *Mona Lisa*, ¿se le imprime que hay una mujer sonriente, o una complicada disposición de evanescentes pigmentos? Cuando se reencuentra con la casa de la abuela, ¿se le imprime que *Es ist etwas unheimliches* o que *c'est une ambiance tout a fait proustienne*? ¿Contiene la disposición del mundo *todas* esas cláusulas-«que», más las que los seres humanos no han tenido tiempo de formular todavía, ya listas y esperando a ser impresas en una sensibilidad visitante?[32]

Valga esto como una estocada rápida para generar malestar entre los que aceptan la prescripción de McDowell. Ahora paso a mitigarlo en quienes se inclinan a aceptar la mía. Dice McDowell que una consecuencia «intolerable» de visiones como la mía es que «cómo sean las cosas […] no puede ser independiente de que la comunidad ratifique el juicio de que son así y así».[33] Mi visión no implica que nuestra comunidad no pueda perder el rumbo, pero sí que cómo sean las cosas no puede ser independiente de un lenguaje en el que se dice que son de esa manera, y que el mundo tiene que aguardar a que lleguemos nosotros con un lenguaje antes de que pueda haber algo ni remotamente parecido al Modo de Ser del Mundo.

Si «cómo son las cosas» significa algo que no sea alguna de las descripciones del mundo que produce la «triangulación» que describe Davidson entre el mundo, los otros y yo, entonces no comprendo esa expresión. La visión de McDowell solo evita el intentar meterse entre el lenguaje y su objeto, como dice Wittgenstein, a base de otorgarle a la disposición del mundo tantos elementos como respuestas lingüísticas hay a sus diversos

segmentos. A mi modo de ver, semejante prodigalidad tiene que crear mucho más malestar que la afirmación de Davidson (citada antes) de que hablar de ajustarse a los hechos (expresión que sintetiza la idea de McDowell de un feliz resultado en forma de «control racional» del mundo sobre la mente) no añade nada inteligible al simple concepto de ser verdadero.

McDowell dice acertadamente que niego que «el sentido todo de la idea de unas normas de investigación es que seguirlas hasta el final mejora nuestras posibilidades de estar en lo cierto sobre "el modo de ser del resto del mundo"».[34] Si uno piensa, como los anti-pragmatistas, que existe semejante modo de ser —que el mundo tiene una disposición que controla racionalmente el pensamiento y no se limita a empujar de un lado a otro a los que piensan—, claro que tiene que parecer que el sentido de las normas de investigación es esa mejora. Pero los pragmatistas como yo (que luchamos por mantener lo más crudo posible nuestro naturalismo) pensamos que su sentido es conseguirnos lo que queremos, y que tratar de descubrir el modo de ser del mundo al margen de cualquier descripción lingüística suya es mala cosa que querer. Los anti-pragmatistas aprueban a los investigadores que quieren ajustarse a los hechos, captar las cosas correctamente, averiguar el modo de ser del mundo, etc. Los pragmatistas no. Estos creen que esos investigadores persiguen quimeras, igual que los agentes que quieren obedecer la Voluntad de Dios. Esas comunidades deberían redescribir sus motivos y metas. No a causa de cómo es el mundo, sino porque, si lo hacen, a la larga sufrirán menos malestar.[35]

* * *

Tanto McDowell como yo queremos ser buenos sellarsianos, pero él me ve a mí como si de la crítica de Sellars al Mito de lo Dado yo infiriera un antirrealismo loco. Cree que de esa forma me he ido de un extremo del balancín al otro, y afirma haber encontrado un modo de terminar con el balanceo entre el Mito de lo Dado, por un lado, y dejar sin más de exigir para la mente constricción racional desde el mundo, por otro. El modo es reconocer que «la receptividad no hace

una contribución separable, ni tan siquiera nocionalmente, a la co-operación [entre receptividad y espontaneidad]».[36] Dice acertadamente que el Mito de lo Dado es una respuesta a la amenaza de que haya insuficiente «constricción externa sobre nuestra actividad en el pensamiento y el juicio empíricos», pero ve «la respuesta davidsoniana» para abandonar ese Mito como una simple batida más del balancín.[37] Por el contrario, yo quiero bloquear el balancín de una vez para siempre del lado donde se abandona la idea misma de constricción racional, no intentar ningún tipo de acción equilibrante.

Cuando discrepamos sobre si mejorar el modelo de Kant de la relación entre mente y mundo o tirarlo a la basura, ¿en qué estamos discrepando McDowell y yo? No, me parece, sobre la mejor forma de reconciliar nuestras diversas intuiciones. El desacuerdo es en realidad sobre qué intuiciones atesorar y de cuáles intentar librarnos. Aquí difiero de Cavell, Putnam y otros que piensan que la filosofía, a la postre, debería dejarlo todo como está, que la tarea de la filosofía es, en palabras que Putnam cita de John Wisdom, completar un viaje desde «lo familiar a lo familiar».[38] Estando de acuerdo con McDowell en que «nuestras angustias filosóficas» en este campo son debidas a «un moderno naturalismo», yo quiero rendirme a ese naturalismo tirando a la basura la imagen de nuestra relación con el mundo que ofrece la distinción esquema-contenido, y en particular las nociones de Kant de espontaneidad y receptividad. Esto significa proponer una visión contraintuitiva.

McDowell dice que su platonismo naturalizado es la idea de que «los dictados de la razón están ahí en cualquier caso, abra uno los ojos a ellos o no; abrírnoslos es lo que hace una educación como es debido».[39] Yo pienso que a lo que nos abre una educación como es debido es a las demandas de una u otra comunidad humana. Dichas comunidades adoptan diferentes normas para investigar dependiendo de lo que quieran sacar de la investigación. Él y yo diferimos sobre si los dictados de algunas comunidades están más en conformidad con los de la razón que los dictados de otras, porque diferimos sobre la utilidad del término «razón». Así que es predecible

que difiramos sobre si hace falta que los filósofos digan mucho más sobre la relación entre razón y naturaleza.[40]

Mientras que McDowell piensa que el que un ser humano a veces esté «por encima de la presión de la necesidad biológica» significa que «hay un poco de realidad objetiva a su alcance»,[41] para mí significa solo que hay placeres que nosotros podemos tener y los cerdos no.[42] Mientras que él cree que «nuestra experiencia misma, en aquel aspecto de su naturaleza que la constituye como experiencia del mundo, participa de una destacada condición del arte, el estar libre de la necesidad de ser útil»,[43] los pragmatistas como yo pensamos que la cognición todavía está uncida a esta última necesidad. Estamos de acuerdo en que el arte está liberado de ella, pero sostenemos que la ciencia no. Nuestro modo de entender la ciencia es utilitarista de principio a fin.

Estas son cuestiones muy vastas y no he hecho más que rascar la superficie de algunos de los profundos desacuerdos que tenemos McDowell y yo. *Mente y mundo* es un libro extraordinariamente rico y sutil, que supone un avance sustancial en la disputa entre los que quieren descartar y los que quieren rehabilitar la imagen de Aristóteles de nuestra relación con el resto de la naturaleza.

---

[1] En cambio, McDowell cree que la afirmación de Davidson de que debemos traducir gran cantidad de los signos y ruidos que algo hace como que son acerca de cosas con las que ese algo está interactuando «no aplaca nuestros temores de que ese cuadro haga que nuestro pensamiento pueda no estar en contacto con el mundo exterior. Justamente, nos da una sensación mareante de que no tenemos un agarre tan firme como pensábamos sobre qué es eso que creemos»; véase John McDowell, *Mind and World*. Cambridge, Massachusetts: Harvard University Press, 1994, p. 17 [hay trad. esp.: *Mente y mundo*, Salamanca, Ediciones Sígueme, 2003]. Aquí McDowell repite un punto que ya ha esgrimido Edward Craig contra Davidson en «Davidson and the Sceptic: The Thumbnail Version» *Analysis* 50, n.º 4 (octubre 1990), pp. 213-214. (El mismo punto lo esgrimen Anthony Brueckner, Crispin Wright y otros contra la explicación de Putnam de cómo lidiar con los cerebros en cubetas.) Los naturalistas crudos, sin embargo, son inmunes al mareo que siente

McDowell. Su agarre sobre qué es eso que creen es exactamente tan firme como su agarre sobre cuál es su entorno, pues no ven modo de identificar un objeto intencional de una creencia (incluso si esa creencia es suya) salvo vinculando la conducta del creyente con el entorno. (Por eso dice Davidson que «la interpretación radical empieza en casa».) Así que consideran que las dudas sobre qué es eso que creemos no están mejor motivadas que las dudas cartesianas sobre cuál es nuestro entorno. Las de este último tipo las despacharían de la forma que se esboza en el argumento de Michael Williams contra Barry Stroud y otros cartesianos; véase Michael Williams, *Unnatural Doubts*. Oxford: Blackwell, 1991.

[2] Por supuesto, lo que Aristóteles quería de la ciencia era muy distinto de lo que quería Galileo. El tipo de destreza que otorga la física corpuscularista para predecir y controlar puede que le hubiera parecido indiferente para la clase de comprensión que debe proporcionar la física.

[3] Davidson, «On the Very Idea of a Conceptual Scheme», en *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 1984, pp. 193-194 [hay trad. esp.: «De la idea misma de un esquema conceptual», en *De la verdad y de la interpretación*, Barcelona, Gedisa, 2001, pp. 189-203].

[4] McDowell, *Mind and World*, p. 15.

[5] *Ibid*., p. 68.

[6] *Ibid*., p. 15.

[7] *Ibid*., p. 14.

[8] Véase *ibid*., p. 67, donde McDowell dice que Davidson no está tentado por «un naturalismo crudo que desertaría en toda regla de este campo de la filosofía, [no está tentado] de negar que la espontaneidad del entendimiento sea *sui generis* de la manera que sugiere el vínculo con la idea de libertad». (Véanse las pp. 108-109 para más detalles sobre lo que hace ser crudo al naturalismo crudo.) Yo creo que la perspectiva de Davidson es tan cruda como puede llegar a serlo el naturalismo. Hasta donde yo veo, todo lo que necesita decir Davidson sobre el carácter sui generis de la mente es que el carácter holístico de la adscripción de estados mentales basta para distinguirlos de estados físicos, y que tal adscripción nos permite hacer una distinción movimiento-conducta, así como una distinción humeana entre libertad y

constricción. No veo qué tiene que ver esto con la espontaneidad ni con lo que McDowell llama la «libertad kantiana». Véanse, sin embargo, las pp. 148-156, donde McDowell argumenta que yo malinterpreto a Davidson. Más adelante me ocupo brevemente de esas páginas.

[9] McDowell, *Mind and World*, p. 24.

[10] *Ibid*., p. 67.

[11] Para Davidson es difícil darle sentido a «contenido conceptual» porque, debo suponer, él considera que tener un concepto es ser capaz de usar un término, una perspectiva nominalista que se ha hecho cada vez más popular desde la aparición de las *Investigaciones filosóficas* y que implica afirmar que ser ciego de nacimiento no impide, *pace* Locke, adquirir el concepto «rojo». Pese a su aprecio por Sellars, McDowell parece disentir de esta idea y coincidir con Locke en que la experiencia visual es un requisito para tener conceptos de color (véase *Mind and World*, p. 167, y también p. 172: «un concepto que se vuelve disponible por la presencia en la experiencia de la muestra original»).

[12] McDowell, *Mind and World*, p. 68.

[13] *Ibid*., p. 142.

[14] *Ibid*., p. 141.

[15] *Ibid*., p. 143.

[16] *Ibid*., p. 160.

[17] *Ibid*., p. 62.

[18] *Ibid*., p. 182.

[19] *Ibid*., p. 178.

[20] *Ibid*., p. 151. Véanse también las pp. 154-155, donde McDowell dice que él y yo compartimos las metas pragmatistas de «desacreditar dualismos y disolver problemas», pero que «el pragmatismo de Rorty mismo está verde si se mide por el rasero que impone su propia explicación de lo que es el pragmatismo». En una nota a pie de página en ese pasaje, dice que no soy

suficientemente consciente de la posibilidad de que Kant y otros «se propusieran tener el derecho a no preocuparse de los problemas [de la filosofía tradicional], no solucionarlos». Yo estoy de acuerdo en que Kant se proponía eso, pero sostengo que creó demasiados problemas nuevos en el proceso de hacer valer tal derecho. Prácticamente lo único que quiero hacer mío de Kant son los pasajes donde da a entender que podemos alternar, como con unas bifocales, entre describirnos a nosotros mismos en términos de causas y en términos de razones, sin que haga falta mediar entre esas dos descripciones (salvo explicando cómo pudo emerger el uso del lenguaje, y por ende el discurso de la razón [*reason-talk*], en el curso de la evolución biológica). Pero McDowell piensa que mi apropiación de esta idea kantiana es simplista y grosera.

21 Davidson, «The Structure and Content of Truth», *Journal of Philosophy* 87 (junio 1960), p. 322.

22 McDowell, *Mind and World*, p. 142n.

23 *Ibid*.

24 *Ibid*., p. 177.

25 *Ibid*., p. 175.

26 *Ibid*., p. 179.

27 *Ibid*., p. 67.

28 *Ibid*., p. 64.

29 *Ibid*., p. 179.

30 *Ibid*., p. 27.

31 *Ibid*., p. 26.

32 Charles Taylor me reprendió una vez por usar la expresión «el propio lenguaje de la Naturaleza» para describir la posición de quienes defienden Nuestras Intuiciones Realistas, dando a entender de ese modo que los realistas eran lo que Taylor llamaba «Platónicos Delirantes» (gente que cree

que «ahí afuera en el mundo hay ya de algún modo un vocabulario»). Prometí no volver a utilizar nunca la expresión. (Véase mi «Taylor on Truth», en *Philosophy in an Age of Pluralism: The Philosophy of Charles Taylor in Question*, ed. de James Tully. Cambridge, Cambridge University Press, 1994, pp. 21-22.) Pero el libro de McDowell me tienta a renegar de mi promesa. Coincido con David Lewis en que «el realismo que reconoce como empresa no trivial la de descubrir la verdad sobre el mundo necesita del realismo tradicional que reconoce identidades y diferencias objetivas, cortes en el mundo, clasificaciones discriminativas que no sean invención nuestra» (Lewis, «Putnam's Paradox», *Australasian Journal of Philosophy* 62, n.º 3 [1984], p. 228). Pero para mí la idea de «clasificaciones que no sean invención nuestra» y de lo que Lewis llama objetos «de élite» es una reducción al absurdo de la de que la verdad es captar correctamente «una realidad independiente», y confirma la tesis de Davidson de que «tan fútil es rechazar el lema de que lo real y lo verdadero es "independiente de nuestras creencias" como aceptarlo» (Davidson, «The Structure and Content of Truth», p. 305. Pienso que la visión de McDowell le fuerza a sumarse a Lewis en lo que para mí es un absurdo.

[33] McDowell, *Mind and World*, p. 93.

[34] *Ibid*., p. 151. Defiendo esta negación en «Is Truth a Goal of Inquiry? Davidson vs. Wright», en mi *Truth and Progress: Philosophical Papers, Vol. 3*. Nueva York: Cambridge University Press, 1998, pp. 19-42 [hay trad. esp.: «¿Es la verdad una meta de la investigación? Donald Davidson *versus* Crispin Wright», en *Verdad y progreso: escritos filosóficos 3*, Barcelona, Ed. Paidós, 2000, pp. 31-61].

[35] Para la tesis de que los beneficios a la larga del pragmatismo compensan la confusión y el desconcierto que causan al principio sus doctrinas, véase mi réplica a la afirmación de John Searle de que Kuhn, Derrida y yo le estamos haciendo daño a nuestra cultura socavando lo que él llama «la Tradición Racionalista Occidental»: «John Searle on Reason and Relativism», en *Truth and Progress*, pp. 63-83 [ed. esp.: «John Searle en torno al realismo y el relativismo», en *Verdad y progreso*, pp. 89-114].

[36] McDowell, *Mind and World*, pp. 8-9. Sospecho que, si postulas dos facultades cuya función es cooperar en una tarea dada y luego privas a una de ellas de contribución separable a esa cooperación, ni tan siquiera nocionalmente, ya no estás postulando *dos* facultades. El monismo hegeliano se viene encima. Yo creo que Hegel tenía razón en que, mientras tengas dos

facultades, acabarás en el escepticismo, y que la única opción es entre una versión hegeliana del monismo y una crudamente naturalista.

[37] *Ibid*., p. 8.

[38] Véase Putnam, *Words and Life*. Cambridge, Massachusetts: Harvard University Press, 1994), p. 300. Defiendo la tesis de que la filosofía es una respuesta al cambio cultural, y que las intuiciones deben cambiar cuando ese cambio se produce, en «Philosophy and the Future» (en *Rorty and Pragmatism*, ed. por Herman Saatkamp. Nashville, Tennessee. Vanderbilt University Press, 1995). Considero que Darwin es la figura crucial del proceso de cambio cultural al que deberían acomodarse los filósofos, y pienso que su obsesión con Kant todavía les impide asimilar a Darwin. Véase mi «Dewey between Hegel and Darwin», en el mismo volumen [reimpreso en *Truth and Progress*. —Eds. / «Dewey, entre Hegel y Darwin», en *Verdad y progreso*, pp. 319-340].

[39] McDowell, *Mind and World*, p. 91.

[40] Véase *ibid*., pp. 153-156.

[41] *Ibid*., p. 116.

[42] McDowell conecta, siguiendo a Gadamer, la distinción cerdo-humano con la de entorno y mundo, y contrapone desfavorablemente a Davidson con Gadamer (véase *ibid*., pp. 114-119, 184-187). Yo creo que se le escapa la gran similitud entre ambos filósofos, que Bjørn Ramberg pone de manifiesto en el capítulo final de su libro *Donald Davidson's Philosophy of Language* (Oxford. Blackwell, 1989). Pero tiene razón en que Davidson vería un *Welt* gadameriano como simplemente el *Umwelt* del cerdo afrontado lingüísticamente.

[43] McDowell, *Mind and World*, p. 119.

CAPÍTULO 16

REDUCCIONISMO VERSUS SEMÁNTICA NEO-WITTGENSTEINIANA

Un reciente libro de Kenneth Taylor, titulado *Truth and Meaning: An Introduction to the Philosophy of Language*, empieza pidiéndole al lector o lectora que se pare a pensar en la pregunta: «¿cómo puede haber enunciados contingentes de identidad, ya que todo es necesariamente idéntico a sí mismo?» Taylor la señala como «el rompecabezas de Frege» y la considera un punto de partida adecuado para reflexionar filosóficamente sobre el lenguaje: el rompecabezas, dice Taylor, es «explicar cómo un enunciado de la forma "a = b" puede diferir en contenido cognitivo de uno de la forma "a = a"».[1] Este es el primero de los ejemplos que pone de fenómenos lingüísticos extraños que han dado origen a la disciplina que denominamos semántica filosófica o filosofía del lenguaje. Otros son la opacidad referencial de los enunciados de creencia, los enunciados verdaderos sobre objetos inexistentes, y algunos más de los ya conocidos sospechosos habituales.

Alguien que llegue al libro de Taylor teniendo fresca la lectura de las *Investigaciones filosóficas* de Wittgenstein, o de «Algunas reflexiones sobre los juegos lingüísticos» de Wilfrid Sellars, o de *Hacerlo explícito* de Brandom, puede que se niegue a pensar que esos sean fenómenos extraños necesitados de explicación. Puede tener la sensación de que Taylor entra en la filosofía del lenguaje con mal pie, con peticiones de principio. Parecida reacción tendrá probablemente quien traiga recién leídos los ensayos últimos de Davidson («A Nice Derangement of Epitaphs» o «The Structure and Content of Truth», por ejemplo). Poco hay del llamado programa Davidson en estos artículos, el intento de encontrar formas extensionalistas de parafrasear enunciados intensionales. En cambio, hablan mucho de la necesidad de no seguir pensando en las

expresiones lingüísticas como representaciones, y de que recordemos las cortapisas de Quine a que puedan existir entidades llamadas significados o intensiones. Cualquiera que dude de esas nociones dudará del uso que hace Taylor de expresiones como «contenido cognitivo» o «contenido representacional» a modo de términos primitivos.

Alguien que ha leído recientemente a Wittgenstein, o a los filósofos del lenguaje neo-wittgensteinianos que he mencionado, quizá reaccione a la descripción de Taylor del «rompecabezas de Frege» recordando un pasaje de las *Investigaciones filosóficas* que dice que no hay más bello ejemplo de oración inútil que «una cosa es idéntica a sí misma».[2] Wittgenstein la compara sarcásticamente con «toda mancha de color ajusta exactamente en su contorno». La evocación puede llevar a la lectora a preguntarse si no será que el único uso de los enunciados de identidad, al menos en los juegos de lenguaje que son su nicho original, es hacer identificaciones contingentes como la del lucero matutino y el vespertino, o la de la mente y el cerebro. La idea de autoidentidad, y a fortiori la de autoidentidad necesaria, es una idea sin la cual todo el mundo parece poder pasarse excepto ciertos filósofos. Si acordamos renegar de ella, no habrá razón para tomarnos en serio el rompecabezas de Frege. Para un wittgensteiniano, el rompecabezas apenas puede servir como paradigma de problema cuya existencia se utiliza para disculpar la de una disciplina que el propio Wittgenstein terminó considerando perfectamente prescindible: la filosofía del lenguaje.

Ese tipo de lectora probablemente se niegue también a aceptar que Taylor apele al test de traducción de Church. Taylor explica que Church criticó la solución de Frege a su rompecabezas invocando dos principios plausibles:

(1) El Principio de Paráfrasis: una oración *S′* es una paráfrasis adecuada de una oración *S* si y solo si lo que expresa *S′* es idéntico a lo que expresa *S*.
(2) El Principio de Traducción: si *T* es una traducción adecuada de *S*, entonces *T* expresa lo que expresa *S*.

Church sacaba punta a esos principios haciendo notar que, si

> «el lucero matutino» denota el mismo objeto que «el lucero vespertino»

fuera una paráfrasis adecuada de

> el lucero matutino es el lucero vespertino,

entonces la traducción al alemán de la segunda oración expresaría justo lo mismo que la traducción de la primera. Pero no es así, porque

> «el lucero matutino» *und* «el lucero vespertino» *bedeutet der gleichen Gegenstand*

no expresa lo mismo que expresa

> *der Morgenstern ist der Abendstern.*[3]

Aquí alguien empapado de Wittgenstein, Sellars, el último Davidson o Brandom, seguramente protestará diciendo que la noción de «expresar lo mismo que» no es más útil que la de autoidentidad. Por supuesto, está de acuerdo en que, a veces, en determinados contextos, dos oraciones diferentes se pueden usar de manera intercambiable. Pero, por razones como las que llevaron a Quine a mofarse de la noción de sinonimia, duda de la idea de que se pueda encontrar una oración que signifique lo mismo que otra *tout court*, sin importar el contexto. Duda, en resumen, de la noción de «encontrar el significado» como distinta a encontrar diversos usos, y de la de algo que las oraciones expresen, o que las haga verdaderas, en todo contexto posible. Inducida por Davidson a dudar de la noción de que las verdades son representaciones de la realidad, cabe que la lectora se pregunte si el uso no-contextual de «expresa lo mismo que» no será una locución tan dudosa y prescindible como «representa el mismo trozo de realidad».

Ítem más, inducida por Brandom a ignorar la pregunta sin contexto «¿qué representa esta oración?» en favor de la pregunta contextual «¿en qué marcador deóntico le anotará los puntos al

emisor de la oración una persona determinada en una situación determinada?», dicha lectora quizá opte por señalar que resulta difícil ver qué tipo de persona, y en qué situación, le daría a la emisión

> (S) el lucero matutino es el lucero vespertino

una puntuación distinta que a

> (S') la cosa a la que los hispanohablantes llaman «el lucero matutino» y la que llaman "el lucero vespertino" son la misma estrella.

o a

> (S") «el lucero matutino» y «el lucero vespertino» denotan el mismo objeto.

Porque la única diferencia de puntuación que ella puede concebir es la que resultara de una situación en la que alguien esté dispuesto a inferir algo de una de esas oraciones que otra persona no inferiría. Semejante situación, sospecha, solo va a darse dentro de un área especial y bastante dudosa de la cultura: la filosofía del lenguaje.

Solo ahí habrá personas que digan que los contenidos representacionales de las emisiones de esas oraciones varían, o que las hacen verdaderas diferentes trozos de realidad, porque solo ahí encontramos usos para expresiones como «denotar», «objeto», «es hecha verdadera por» y «expresa lo mismo que». Dentro de esa área de la cultura no se admitirán apelaciones al uso común, por supuesto, como no se habrían admitido en las discusiones del siglo XVII en torno a si usar o no, y cómo, expresiones como «movimiento natural», «movimiento violento», «cantidad de movimiento», «inercia» y similares. Así que, de nuevo, la pregunta que plantean wittgensteinianos, brandomianos y demás es: ¿la filosofía del lenguaje es un área de la cultura donde el lenguaje sencillamente se ha ido de vacaciones, o es un área en la que se abordan problemas reales con ayuda de un vocabulario convenientemente técnico y separado del sentido común?

Esa pregunta se reduce a esta otra: ¿de qué creemos que nos sirve la filosofía del lenguaje? ¿Hay, como piensa Taylor, fenómenos extraños que necesitan explicación, como en los tiempos en que la gente debatía sobre la utilidad relativa de los términos «movimiento natural» e «inercia», o hay, como dicen displicentes los terapeutas wittgensteinianos, meros pseudoproblemas creados por una imagen de la que estábamos cautivos?

Plantear la alternativa de esa manera da a entender que podemos ver como filósofos terapéuticos, en vez de sistemáticos y constructivos, no solo a Wittgenstein sino también a Sellars, Davidson y Brandom. En esa insinuación, implausible a primera vista, hay un punto de verdad. Porque los cuatro filósofos que acabo de mencionar tienen un blanco común. Los cuatro intentar libertarnos del cautiverio a que somete la misma imagen. Que algunos lo hagan ofreciéndonos vastas estructuras de argumentos interconectados, y no con aforismos breves y tonificantes, no debería oscurecer el hecho de que todos ellos ven como irreales ciertos problemas que Taylor, David Lewis, Saul Kripke, David Kaplan y muchos otros consideran reales.

Esos son los problemas que se espera que resuelvan nociones como «intensiones» y «significados», problemas creados por la tesis de que los mismos objetos y estados de cosas pueden ser descritos de muchas maneras diferentes. La idea de Frege es que hay unas cosas llamadas intensiones cuyo grano es más fino que el de las extensiones representadas, y que la filosofía del lenguaje busca correlacionar esas intensiones con el uso de expresiones lingüísticas, mostrando así de qué manera logra el lenguaje funcionar como un sistema representacional.

Dice Taylor, en tono desconcertado, que en la perspectiva de Davidson «los significados no desempeñan ningún papel en la teoría del significado».[4] La misma observación podría hacerse, mutatis mutandis, para Wittgenstein, Sellars y Brandom. En la medida en que esos filósofos tienen teorías del significado, no son de la clase que puede usarse para averiguar qué significa una expresión lingüística dada. Salvo Davidson, no se dedican a parafrasear

expresiones buscando otras equivalentes y más perspicuas, e incluso, en el caso de Davidson, esa paráfrasis es subsidiaria del proyecto más amplio de explicar holísticamente, en vez de piedra a piedra, cómo se relacionan las expresiones lingüísticas con entidades no lingüísticas. Taylor ve el holismo semántico de Davidson —que define como la tesis de que «los significados de las expresiones de un lenguaje dependen de la totalidad de relaciones que esas expresiones mantienen entre sí»— como un modo de hacer que las nociones semánticas sean «respetables en términos fisicistas aun cuando no pueda haber una reducción fisicista de los primitivos semánticos».[5]

El descreimiento hacia los proyectos reduccionistas se intensifica cuando uno lee a Wittgenstein y empieza a pensar que toda expresión lingüística tiene un significado del todo respetable meramente en virtud de tener un uso, y que cualquier expresión tiene un uso si tú se lo das. Pese a epítetos de dudosa consistencia, como «sinsentido», empleados por el propio Wittgenstein para caracterizar expresiones como «una cosa es idéntica a sí misma», los lectores de las *Investigaciones filosóficas* son reacios a decir que algunas expresiones efectivamente en uso carecen de «un significado claro». Porque, si no se tiene en mente algún programa reduccionista específico, la intraducibilidad de esas expresiones a un vocabulario alternativo en nada demuestra ausencia de significado o ausencia de claridad.

Aunque, en las *Investigaciones*, Wittgenstein sigue hablando a veces de detectar confusiones conceptuales, en la mayoría de casos se contenta con permitir que mil juegos de lenguaje se jueguen[6], sin dar a entender que necesiten supervisión filosófica. Davidson y Brandom no hablan nunca de confusiones conceptuales o sinsentidos. Es así porque ambos siguen a Quine en repudiar la idea misma de unos objetos estables que se ven más claramente y se comprenden mejor cuando se los describe con un conjunto de términos, los *analizantes*, que cuando se los describe con otro, los *analizados*. El precio que pagó Quine por preservar la filosofía del lenguaje como tema serio, según lo expresó Davidson, fue sembrar

la duda tanto sobre el reduccionismo, por el que él mismo sentía cariño, como sobre la idea carnapiana de que la filosofía del lenguaje podía ser un instrumento de crítica de la cultura.

La filosofía del lenguaje anglófona contemporánea, tal como yo la veo, se ha escindido en dos campamentos. En el primero o neo-wittgensteiniano, están aquellos cuyos filósofos del lenguaje favoritos son Davidson y Brandom. En el segundo o reduccionista, los que ponen entre sus favoritos a David Lewis, Saul Kripke, David Kaplan y John Perry. Estos últimos sintonizan de manera característica, bien con el tipo de metafísica fisicista que Lewis comparte con Frank Jackson y muchos otros filósofos de Australasia, bien con el intento que hacen Dretske, Millikan y Gibbard de basar la semántica y la epistemología en la biología evolucionista, bien con ambas cosas. Si Russell y Carnap volvieran de la tumba, encontrarían más inteligible el trabajo en este segundo campamento que en el primero. Porque el del segundo continúa la idea de «la filosofía de la ciencia unificada» de una forma que no hace el primero.

La idea de que la filosofía puede volverse científica llevando hasta el final el tipo de programa de investigación reduccionista que concibió Carnap va de la mano con la tesis de que una de las funciones de la filosofía es depurar la cultura librándola de elementos «acientíficos». Esa tesis ya no se presenta en el tono más bien tosco e irrespetuoso de Ayer y Popper. Pero la insatisfacción que sienten muchos filósofos del segundo campamento con el trabajo que realizan los del primero tiene muchísimo que ver con la necesidad de mantener vivo algo parecido a la fervorosa cruzada de Carnap y Popper, con la necesidad de fabricar armas que mantengan a raya a los filósofos política y espiritualmente peligrosos (Hegel y Heidegger, por ejemplo).

Otra fuente de su insatisfacción es que cuesta concebir un programa de investigación en semántica wittgensteiniano o brandomiano. Ninguno de los dos ofrece una bolsa de problemas irresueltos para que les hinquen el diente los graduados en sus tesis. En la década de 1980, parecía como si Davidson tuviera un programa así que proponer, y ese espejismo tuvo bastante que ver

con la entusiasta acogida que recibieron sus *Locke Lectures* en Oxford. Pero, en los años subsiguientes, parece haber perdido interés en tallar a escoplo las expresiones lingüísticas en hormas extensionales. Su trabajo reciente parece tan lejos como el de Brandom de cualquier intento por contestar a la pregunta «¿qué queda ahora por hacer?»

I
Brandom como naturalista no-reduccionista

Valga lo anterior como rápida revista a la escisión que yo tenía en mente al elegir el título de este artículo. Quiero volver ahora a la cuestión de la respetabilidad de las nociones modales, y considerar la tesis de Brandom de que existe solo una buena razón para que la hayan reencontrado: que nos hemos dado cuenta de que todo uso de cualquier concepto presupone la capacidad de blandir un vocabulario normativo.

Cuando captamos eso, piensa Brandom, comprendemos que no podemos ni siquiera usar «¡rojo aquí ahora!» como emisión significativa —esto es, como una jugada en lo que Brandom llama «el juego de dar y pedir razones»— a menos que entendamos y aceptemos expresiones que formulan normas para el uso de las palabras que la componen y para la expresión como un todo. Como hasta el reduccionista más estricto estaría de acuerdo en que un vocabulario normativo proporciona recursos suficientes para «clarificar» contrafácticos y locuciones modales, si tienes ese vocabulario, tienes todo. Podemos, por ejemplo, definir ser nomológico en términos de ser proyectable, y explicar la proyectabilidad en términos de locuciones como «si proyectas ese predicado, te molemos a palos o te retiramos la licencia docente, y a lo mejor las dos cosas».

La amenaza del dualismo cartesiano, que fue lo que movió inicialmente a preocuparse por las oraciones de creencia, ahora se desvanece. Porque, como dice Brandom, hemos saltado «de un dualismo a grandes rasgos cartesiano de lo mental y lo físico a un dualismo a grandes rasgos kantiano de lo normativo y lo fáctico».[7]

Pero este segundo dualismo es perfectamente compatible con la explicación naturalista de Darwin de cómo llegó al mundo la intencionalidad. Porque una norma es solo un hecho sociológico —una regularidad de conducta— visto desde dentro. El sociólogo nos dice que, como cuestión de hecho, un determinado grupo te golpea con palos o te retira el título bajo ciertas condiciones. Los miembros de ese grupo nos hablan de las normas a las que están sujetos, pero, como dice el reduccionista, en ese estar sujetos no hay «nada más» que la prevalencia de las regularidades de las que informaba el sociólogo. De manera análoga, un hecho es solo una norma vista desde fuera: el hecho, por ejemplo, de que los miembros de un cierto grupo se sienten sujetos a usar la expresión «¡rojo aquí ahora!» solo bajo ciertas condiciones, so pena de ser molidos a palos o ver canceladas sus cualificaciones académicas.

Aquí la jugada decisiva es interpretar «en una norma no hay más que una regularidad», no como la tesis de que un enunciado que formula una norma se puede remplazar por otro que informa de una regularidad, sino como la tesis de que la explicación causal de cómo llegó a existir la norma es la misma que la de cómo llegó a existir la regularidad. La diferencia entre los dos campamentos de la filosofía del lenguaje contemporánea se reduce a la cuestión de si esa diferencia marca diferencia alguna: si la mera aserción de una superveniencia puede remplazar a una larga serie de análisis del significado y si lo puede hacer sin pérdida de comprensión filosófica.

Esa pregunta sugiere otra más general: ¿cuánto respaldo necesita la aserción de que algo es superveniente a lo físico? Si decimos que el problema central de la filosofía contemporánea es evitar caer en cualquier tipo de dualismo, y de ese modo animar a quienes creen en ectoplasmas a salir de su escondite, puede parecer que los filósofos contemporáneos están predicando a los ya convertidos. Porque la mayoría de los que leen libros, incluidas muchas personas devotamente religiosas, hace mucho que se convencieron de la verdad del fisicismo entendido como la tesis de que, siguiendo la formulación de Frank Jackson, «cualquier mundo que sea un duplicado físico mínimo del nuestro es un duplicado *simpliciter* de

nuestro mundo»[8] y, por consiguiente, un duplicado psicológico, un duplicado sociológico, un mundo que contiene todas las normas que hay en nuestro mundo y solo ellas, y un mundo en el que solo son necesarios o posibles los mismos estados de cosas y solo ellos. Entre la alta cultura de Occidente, tener miedo a los fantasmas y agradecer que nos lo quiten no es que sea la cosa más frecuente del mundo.

Más adelante volveré sobre esta cuestión de la utilidad de la filosofía del lenguaje para la cultura en su conjunto. De momento solo quiero hacer notar que la razón de Brandom para considerar respetables los términos modales no puede ser la razón de un reduccionista, y por tanto su explicación de por qué esos conceptos son ahora tan ampliamente aceptados falla como explicación sociológica.

Los reduccionistas como Jackson quieren saber qué estados de cosas físicos hacen verdaderos a los enunciados normativos o modales. Como dice Jackson, «la metafísica seria exige que respondamos cuándo las cuestiones descritas en un vocabulario son hechas verdaderas por cuestiones descritas en otro».[9] Que Brandom nos asegure que el discurso normativo es inevitable, en el sentido de que hay que emplearlo si es que se quiere emplear cualquier discurso, y que el discurso modal se puede rebajar a discurso normativo, no ayuda en nada a responder a eso.

La estrategia del propio Brandom no puede ayudar a lo que él llama, citando a Dretske, el proyecto reduccionista de hacer caldo intencional a partir de huesos no-intencionales. Brandom nos dice qué no podemos hacer sin hacer algo más, qué jugadas inferenciales tenemos que ser capaces de hacer si vamos a blandir este o aquel concepto. Pero decirnos qué inferencias y conceptos son parasitarios de qué otros no ayuda a identificar los significados singulares de las expresiones, o a delimitar qué expresa una expresión dada o qué hace verdadera a una determinada oración verdadera.

Doy por hecho que Brandom nunca ha dudado de que un duplicado físico de nuestro mundo contendría todas las normas que hay en él y solo ellas, pero no creo que le haya atraído nunca lo que

Jackson llama el «análisis conceptual, entendido como el oficio de encontrar condiciones necesarias y suficientes mediante el método de los casos posibles», es decir, el método de averiguar qué nos dice la intuición que diríamos en diversas situaciones fantásticas e impredecibles.[10] Tal método lo ilustra admirablemente, piensa Jackson, el descubrimiento de Putnam de que la sustancia acuosa de la Tierra Gemela no es un tipo diferente de agua, sino que no es agua en absoluto, y el de Gettier de que la creencia verdadera justificada es insuficiente para el conocimiento.

Desde el punto de vista de Brandom, si le entiendo bien, no hay nada de interés filosófico particular que pudiera depender de que coincidamos con Gettier o Putnam o, por el contrario, tengamos intuiciones —como Jackson admite que les ocurre a algunas personas— que nos lleven a usar «conocimiento» o «agua» de modo distinto a como esos filósofos creen que debe usarse. Solo quienes piensan que existen cosas tales como intensiones que determinan los límites de las extensiones, al modo en que los moldes para galletas determinan los límites de las galletas, podrían interesarse en la cuestión de qué intuiciones prevalecen. Si no crees que existan esas cosas, te tomarás las controversias sobre si en la Tierra Gemela no hay agua, o sí la hay pero de un tipo diferente, como algo igual de inane para establecer la verdad del fisicismo que las controversias sobre cuántos sacramentos instauró Cristo. Considerarás que, una vez se han hecho explícitos los dos diferentes conjuntos de inferencias desde «XYZ es el líquido común transparente bebestible de la Tierra Gemela» que hacen las personas con intuiciones contrapuestas, y que una vez se han hecho igualmente explícitos los dos conjuntos diferentes de inferencias que extraen luteranos y católicos de la oración «esto es la ordenación de un sacerdote», nada puede depender de la cuestión ulterior de si, o cómo, esas dos oraciones son hechas verdaderas por qué estado de cosas físicamente descrito.

Presumiblemente, Jackson consideraría la adhesión al fisicismo contenida en el libro de Brandom como lo que él llama un «acto de fe», salvo que Brandom esté dispuesto a pasar luego a picar piedra,

que es hacer análisis conceptual.[11] Sin embargo, es obvio que el libro de Brandom no se limita a aseverar de forma indolente y como ya sabida que lo normativo superviene a lo físico. Es más como un recorrido por la secuencia de sucesos en que consistió el llegar a ser de las normas, la intencionalidad, la modalidad y todo lo demás que para los filósofos pudiera ser evidencia de que los fantasmas han estado interviniendo. Pero esa descripción también es engañosa. Porque Brandom no está dándonos una cronología de sucesos, sino más bien una explicación de cómo todo eso no habría llegado a ser a menos que otras determinadas cosas —la referencia anafórica, los términos singulares y las comillas de advertencia[12], por ejemplo— hubieran emergido también. Se parece más a un biólogo diciendo «antes de poder tener lo que llamamos visión, había que tener tejidos de este tipo, y de este, y de este», que al mismo biólogo diciendo «primero ocurrió esto, y después esto otro, y después, por primera vez en la historia del planeta, exactamente a las 2:42 PM, un organismo vio algo».

Cualquier biólogo que dijera algo por el estilo estaría sobrepasando tanto sus competencias como nuestro grado de interés en el asunto. Lo único que realmente queremos saber es que él entiende cómo se las arregló la evolución para hacer el truco. Solo queremos los detalles que basten para ayudarnos a captar qué hizo falta para que el resultado se produjera. Eso es lo que Brandom nos da. Él hace una pregunta que rara vez los filósofos se han hecho: ¿cuántas jugadas y de qué tipo han de tener lugar antes de que podamos decir «¡eh, ahora esos organismos están dando razones!»

En su admirable librito *Rationality*, Jonathan Bennett se preguntaba si las abejas contaban como racionales. Concluía que no, aunque solo fuera porque su llamado lenguaje no tenía cuantificadores ni un operador de negación.[13] Brandom empieza remontando la senda que abrió Bennett, pero la extiende mucho más allá de lo que este previó. Su respuesta a la pregunta «¿qué hace falta para ser racional?» es, en un sentido, muy común: para ser racional tienes que usar la lógica. Pero su manera de explicar exactamente por qué cualquier especie que vaya a usar patrones de

signos y sonidos como pegamento entre grandes grupos de sus miembros para emprender proyectos cooperativos tendrá que ser capaz de formular y suscribir todas las inferencias que aparecen en los libros de lógica elemental, es verdaderamente nueva.

La descripción que hace Brandom de su propia empresa es que asegura que «el *advenimiento* de la intencionalidad no quede como algo aparentemente mágico o misterioso».[14] Todo aquel dispuesto a considerar lo psicológico como superveniente a lo físico ya habrá concluido que no es ni una cosa ni la otra. Pero oír que el advenimiento de la lógica es simultáneo al de lo psicológico es sorprendente. Pues tendemos a pensar que la lógica se relaciona con el pensamiento como la forma con la materia: extrayendo orden del caos, constriñendo impulsos indisciplinados. Asusta oír, como le oímos a Brandom, que tener creencias inconsistentes no es irracional, es el sino común de la humanidad. También asusta oír que no existe tal cosa como un pensamiento irracional, en tanto que distinto a un pensamiento que desafía las expectativas razonables de una comunidad humana.

Una forma de describir el cambio de clima filosófico que Brandom desea provocar en nuestra comprensión de la relación entre lógica y psicología es indicando que la platónica división tripartita del alma —que Descartes aceptó sin modificaciones sustanciales— nos inclinaba a concebir que un ser humano que actúa responsablemente era alguien que había puesto las partes más bajas y peores de su alma bajo la constricción de la más alta y mejor. En cambio, Brandom nos dice que la única constricción es la que recibe el organismo individual desde el grupo circundante de organismos. No tenemos responsabilidades, ni hacia la ley moral ni hacia las de la lógica, que no se reduzcan a nuestra responsabilidad de estar a la altura de la confianza que depositan en nosotros los seres humanos que tenemos delante. Brandom está continuando la línea wittgensteiniana/ryleana de desacreditar la idea platónica/cartesiana de la mente como un componente extra, añadido a nuestra naturaleza animal, al decir que lo que nos hace humanos son las

relaciones de confianza que nos unen a otros miembros de nuestra especie.

A Russell, Husserl y Carnap, esta forma de pensar seguramente les habría sonado como una regresión al psicologismo del que se supone que Frege nos ayudó a salvarnos. Pero el sociologismo de Brandom también se puede ver como una forma de rescatarnos del psicologismo, o por lo menos de la modalidad cartesiana de psicologismo que interpreta la mente como algo que todo ser humano puede tener totalmente por sí solo. El objeto rescatado es el mismo en ambas iniciativas filosóficas: es el individuo errado que necesita un sentido más grande de responsabilidad. Pero, en Brandom, la fuente del rescate no es no-empírica y ahistórica, sino un grupo de personas históricamente contingente nada más.

Brandom puede responder a la pregunta retórica de Wittgenstein «¿por qué pensábamos que la lógica era algo sublime?» diciendo «porque creíamos que podíamos apelar desde las comunidades locales a algo universal, y porque confundimos universalidad con existencia necesaria». La lógica ciertamente es universal, pero solo porque está implícita en cualquier cosa a la que llamemos práctica lingüística. Una vez que seguimos a Quine y Davidson en ver que para la traducción hace falta caridad, podemos ver que nunca vamos a cruzarnos con algo a lo que pueda llamarse un uso del lenguaje sistemáticamente ilógico o sistemáticamente irracional. Lo peor con lo que nos podemos encontrar es con un grupo de compañeros conversacionales que discrepan de nosotros en una gama muy grande de temas.

El sociologismo que recomienda Brandom como forma de salvar a la lógica del psicologismo equivale a decir que debemos obedecer las reglas de la lógica, no porque se alcen augustas por encima del mundo del tiempo y del cambio, sino porque hacerlo nos permite cooperar con otras personas. La constricción de la lógica no es la que lo permanente y estable ejerce sobre lo transitorio, no el dominio que la forma ejerce sobre la materia, sino la constricción que ejerce el grupo sobre el individuo potencialmente no cooperativo.

II
La semántica brandomiana como terapia y como historia natural

Supongamos que definimos la semántica como la descripción pormenorizada de los significados de diferentes emisiones lingüísticas, y a continuación definimos la semántica brandomiana como el tratar el significado de una emisión como la puntuación deóntica que se le anota. Tendremos que decir entonces que, para los brandomianos, no habrá solo un significado de una emisión, sino tantos como puntuadores deónticos diferenciables de modo relevante. No obstante, a menos que exista una forma de privilegiar a unos puntuadores sobre otros, no habrá manera de utilizar la semántica para criticar las emisiones de nadie.

Hasta donde yo sé, Brandom no nos da esa forma. Nunca interviene en disputas entre puntuadores para aportar más claridad. Porque su postura le exige dejar la puntuación deóntica a la comunidad lingüística relevante, comunidad que puede ser una minúscula cultura experta o el público como un todo. El filósofo brandomiano del lenguaje coincide con Quine en que debemos abandonar la idea del significado como «museo» —tomar los significados como un conjunto de objetos duraderos e invariables etiquetados mediante expresiones lingüísticas—, pero rompe con Quine en lo que hace a la cuestión de si la filosofía, en cuanto que distinta de las artes y las ciencias, puede mostrar que es mejor usar algunas expresiones lingüísticas y no otras. Para Quine, como para Carnap, el filósofo del lenguaje todavía puede serle útil a la crítica de la cultura. Para Brandom, como para Davidson, no puede. Una vez que dejas el mito del museo, dicen estos últimos filósofos, no estás en situación de decirle a nadie de qué está hablando «realmente» cuando usa una determinada expresión. Ni estás en situación de decirle que su uso del lenguaje está confundido. Como mucho, podrías decirle que es insólito o improductivo.

Entonces, ¿de qué vale un filósofo del lenguaje que no puede diagnosticar la falta de claridad o la confusión? Bueno, algo que puede hacer es ayudar a la cultura más amplia a comprender por

qué no existen caminos que lleven desde una explicación de cómo funciona la puntuación deóntica a una crítica de las prácticas de puntuación contemporáneas. Eso es algo que necesita hacerse, porque si sabes cómo opera la puntuación deóntica, sabes todo lo que se sabrá jamás sobre la intencionalidad y, por consiguiente, sobre temas tradicionalmente tan oscuros y profundos como el de «la naturaleza de la mente». Si la noción de crítica de la cultura puede ensancharse para incluir la actividad de explicarle a la gente por qué la filosofía no va a hacer algunas de las cosas que a veces se esperan de ella, entonces Brandom, Davidson y Wittgenstein pueden contar como críticos de la cultura.

Intentar ir desde la semántica brandomiana a la crítica cultural en el sentido más restringido —el tipo de crítica de la cultura que creían estar practicando Carnap y Popper— sería como intentar ir desde una comprensión de qué hace falta para que una práctica social cuente como un juego de naipes a una crítica de los diferentes juegos incluidos en el Hoyle[15]. Hacer más explícito lo que el Hoyle da por sentado no le ayuda a uno a decidir si jugar al póquer o al bridge. Hacer explícito por qué necesitas anáforas para poder tener términos singulares no te ayuda a decidir qué términos singulares usar o qué inferencias hacer desde cualquier emisión concreta que contenga los que tú usas.

No obstante, el libro de Brandom solo puede tener el deseable efecto terapéutico que yo le estoy atribuyendo en quienes acepten la tesis de que nada hay de profundo o de oscuro en la mente salvo la irreductibilidad de lo intencional, y la tesis ulterior de que esta última se reduce a la irreductibilidad de lo normativo. Nadie que no esté dispuesto a ver esas tesis como sugerencias siquiera plausibles se va a molestar en leer *Hacerlo explícito*. (De ahí que leerlo no les reporte beneficio alguno a Searle o Nagel, por ejemplo.) A las personas no dispuestas a eso no las impresionará el argumento en favor de la ineliminabilidad de lo normativo que resumí antes, y les seguirá pareciendo que el discurso modal necesita algo del estilo de una semántica de los mundos posibles para poder hacerlo respetable.

En este punto es justo preguntar: a los que ya habíamos aceptado ambas tesis, ¿qué nos dice Brandom que no supiéramos antes? ¿Qué añade exactamente su neologismo «puntuación deóntica» a la noción familiar de juego de lenguaje? La respuesta más simple quizá sea: nos muestra cuáles son las constricciones sobre el uso de signos y ruidos que permiten que se diga de nosotros que estamos *razonando* y no despotricando en formas consuetudinarias y aceptadas. La razón de que *Hacerlo explícito* sean seiscientas páginas de picar piedra, y no doscientas de aforismos desacreditadores, es que nunca nadie había combinado antes las virtudes de las *Investigaciones filosóficas* y del *Tractatus Logico-Philosophicus*. Nadie ha explicado por qué los lenguajes tienen que ser en muy buena medida como son si tienen que servir a los propósitos a los que sirven, aparte de los filósofos que piensan que el principal propósito de los lenguajes es representar estados de partículas físicas elementales con ayuda de abreviaturas cómodas.

Para Brandom, en cambio, el propósito principal de los lenguajes es capacitarnos para entablar proyectos cooperativos de una complejidad significativamente mayor que los de las abejas y los castores. Los lenguajes son como son porque, por ejemplo, los usuarios del lenguaje han de tener anáforas, comillas de advertencia y términos singulares para poder lograr tal cosa. Resulta que no serías capaz de dar y pedir razones si no hubieras dominado esas jugadas básicas, igual que no serías capaz de construir una catapulta o una computadora si no hubieras dominado la palanca, la polea y el resto de las llamadas máquinas simples. La diferencia entre la concepción de Brandom de la lógica y la de Russell, Carnap y Husserl es la diferencia entre algo a lo que se debe respeto —en especial si queremos hacer lo que más le cuadra al ser humano, representar la realidad con exactitud— y algo que tenemos que usar para conseguir montones de cosas que queremos conseguir. Esa diferencia corre en paralelo con la que hay entre hacer que las representaciones se desprendan de inferencias y tratar las inferencias como un modo de hacer malabares con representaciones.

[1] Kenneth Taylor, *Truth and Meaning: An Introduction to Philosophy of Language*. Malden, Massachusetts: Blackwell, 1998, p. 2.

[2] Ludwig Wittgenstein, *Philosophical Investigations*, trad. De G. E. M. Anscombe. Nueva York: Macmillan, 1968, § 216 [N. del T.: seguimos, con una mínima rectificación, la traducción al castellano de A. García Suárez y U. Moulines, *Investigaciones filosóficas*, Barcelona, UNAM / Editorial Crítica, 1988. La oración entrecomillada de la frase siguiente es también una cita literal del mismo parágrafo de las *Investigaciones*.]

[3] Taylor, *Truth and Meaning*, pp. 4-5.

[4] *Ibid*., p. 147.

[5] *Ibid*., pp. 175-176.

[6] N. del T.: la expresión remeda el célebre llamamiento de Mao Zedong: «permitir que cien flores florezcan…» (convertidas a menudo en «mil flores» en las referencias occidentales).

[7] Robert R. Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Massachusetts: Harvard University Press, 1994, p. 623 [hay trad. esp.: *Hacerlo explícito: razonamiento, representación y compromiso discursivo*, Barcelona, Ed. Herder, 2009].

[8] Frank Jackson, *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Nueva York: Oxford University Press, 1998, p. 12.

[9] *Ibid*., p. 41.

[10] *Ibid*., p. 60.

[11] *Ibid*., p. 29.

[12] N. del T.: *scare quotes* (literalmente, citas o comillas «del miedo»). Es el uso que se hace de ese signo ortográfico cuando el emisor desea distanciarse de la expresión entrecomillada, bien para indicar ironía, bien porque le parece incorrecta o inapropiada, bien porque no suscribe todas sus connotaciones, etc. En la comunicación oral, decimos «entre comillas» a continuación de la

expresión (o las gestualizamos con los dedos) justamente para reproducir ese mismo efecto de la comunicación por escrito.

[13] Jonathan Bennett, *Rationality: An Essay Towards an Analysis*. Indianapolis, Indiana: Hackett, 1989.

[14] Robert Brandom, «Modality, Normativity, and Intentionality», *Philosophy and Phenomenological Research* 63, n.º 3 (2001), p. 607.

[15] N. del T.: el inglés Edmond Hoyle (1672-1769) fijó las reglas de diversos juegos, sobre todo de naipes, en unos folletos que rápidamente pasaron a convertirse en fuente de autoridad, y hoy muchos compendios de reglas se publican todavía con la palabra «Hoyle» como sello de garantía aunque no procedan del original. En inglés coloquial, la expresión «de acuerdo con Hoyle» viene a significar algo parecido a «conforme mandan los cánones», en relación con cualquier materia.

CAPÍTULO 17

OBSERVACIONES SOBRE NISHIDA Y NISHITANI

Mi colega Norihide Suto me sugirió amablemente unos pocos textos de Kitarō Nishida y Keiji Nishitani que yo pudiera leer para hacerme alguna idea de la tradición intelectual de la Universidad Otani. Le estoy muy agradecido por ello. Como es obvio, mi conocimiento de estos autores es extremadamente superficial y soy casi del todo incapaz de relacionar los pocos textos suyos que he leído con el trasfondo budista que hay en su pensamiento. No obstante, leyéndolos he pensado que unas cuantas observaciones sobre la relación entre pragmatismo e idealismo absoluto servirían de útil prefacio a las cuatro conferencias sobre el pragmatismo que voy a impartir.

Muchos de los argumentos que Nishida y Nishitani encuentran en autores occidentales como Leibniz, Spinoza y Hegel, y que suscriben con entusiasmo, son variaciones sobre el siguiente conjunto de inferencias:

1. Conocer un objeto no es obtener una imagen mental exacta de él aisladamente de todo lo demás, sino, al contrario, ponerlo en relación con otras cosas.
2. De manera que conocer un objeto completamente, adecuadamente, como verdaderamente es, sería ponerlo en relación con todas las demás cosas.
3. De manera que el conocimiento de cualquier objeto, incluido un yo humano, exige el conocimiento de todas las cosas, interrelacionadas en un único sistema coherente. En un sentido, solo hay un objeto que conocer. La pluralidad es una ilusión.
4. Semejante conocimiento lo sería de lo que Spinoza llamaba «Dios o la Naturaleza». Tener tal conocimiento sería experimentar lo que Hegel concebía como la unión final de

sujeto y objeto, Hombre y Dios, la Autoconciencia Absoluta del Espíritu.

Llamemos a esta cadena de inferencias el Argumento desde el Holismo al Monismo. El holismo, en este sentido, es la doctrina de que debemos rechazar el atomismo de Aristóteles y Locke, por ejemplo (que consideraban suficiente la intuición intelectual o sensorial para captar la entera naturaleza de determinadas entidades). En vez de ello, debemos pensar que una cosa no es más que sus relaciones con todo lo demás.

Este argumento se halla en muchos filósofos anglófonos de finales del siglo XIX, como T. H. Green, F. H. Bradley y Josiah Royce. De ellos, el partidario del idealismo absoluto que mejor conocían los fundadores del pragmatismo era Royce, cuyos libros (como *The Religious Aspect of Philosophy* y *The World and the Individual*) eran muy populares en Estados Unidos a comienzos del siglo XX. Royce creía, igual que Nishida, que existía una vía puramente filosófica, argumentativa, hacia el mismo pensamiento que hallamos en muchos escritores religiosos: que el supremo autoconocimiento consiste en tomar conciencia de la identidad del yo con Dios. Como Nishida, Royce concebía la vida divina como «un ahora eterno», y que el autoconocimiento comienza cuando comprendemos que la existencia plural (por ejemplo, la espaciotemporal) es una noción internamente inconsistente. Bien podría haber suscrito las palabras de Nishida: «el yo solo es auténtico cuando conoce su propia muerte eterna. Una vez que hemos captado este hecho, ya nos encontramos existiendo en la vida eterna».[1]

James y Dewey conocían personalmente a Royce y estaban totalmente familiarizados con sus tesis y argumentos. Pero veían en el Argumento desde el Holismo al Idealismo Absoluto una *reductio ad absurdum* de la tradición metafísica, intelectualista. Consideraban que el idealismo absoluto era una vía muerta.

El pragmatismo es una reacción rotunda contra Hegel, contra el idealismo absoluto y contra la metafísica, reacción que fue posible debido a la necesidad que sentían filósofos como James, Dewey y Nietzsche de asimilar la teoría de la evolución de Darwin. Esa

reacción consistió en decir que el conocimiento —en el sentido de representación adecuada de, o unión con, un objeto— no es la meta de la vida humana. Lo que llamamos la búsqueda de conocimiento debería verse como la búsqueda de una acción exitosa, donde «exitosa» significa «que conduce a una mayor felicidad humana». Esto quiere decir que debemos abandonar la idea de un conocimiento «completo» o «adecuado» (ya sea del yo o de cualquier otra cosa) como la meta, o incluso como una de las metas, de la vida humana. En lugar de ello, deberíamos vernos como animales listos.

Desde este punto de vista, la meta de nuestras investigaciones científicas, políticas y de otro tipo es simplemente captar las relaciones entre aquellos rasgos de nuestra situación que queremos cambiar y otros que no queremos que cambien, de forma que podamos hacernos con herramientas eficaces para mejorar esa situación. No debemos decir, como dirían Nishida y Spinoza, que «la existencia racional nunca puede contener en sí misma autocontradicciones», sino más bien lo contrario: que la existencia racional es precisamente el proceso de lidiar con contradicciones, no para llegar a un estado libre de contradicción, sino para resolver los problemas prácticos que esas contradicciones crean.

Podemos interpretar que Dewey contrarresta el Argumento desde el Holismo al Idealismo con el siguiente Argumento desde el Holismo al Pragmatismo:

1. Conocer completamente un objeto sería, de manera imposible, conocer sus relaciones con todos los demás objetos y, por consiguiente, saberlo todo acerca de todo.
2. Puesto que ese es un ideal imposible, deberíamos abandonar la idea misma de «conocer completamente un objeto» (y a fortiori la de conocer el yo completamente, o adecuadamente, o como verdaderamente es).
3. Debemos dejar de concebir el conocimiento como una unión de sujeto y objeto, o la adecuada representación del objeto por el sujeto, y en su lugar concebirlo como una destreza en manipular objetos, incluida la de redescribirlos.

4. Deberíamos dejar de concebir como tarea del ser humano la de conocerse a sí mismo como verdaderamente es (concepción que, en Occidente, se remonta hasta los griegos), y en su lugar concebir que su tarea es cooperar socialmente con otros seres humanos para construir un mundo mejor; más concretamente, para construir una utopía social igualitaria.
5. Cambiar de esa manera el sentido que tenemos de cuál es el propósito de la vida humana tendrá el efecto de secularizar la cultura. Las personas pasarán a verse a sí mismas como animales listos nada más, no como mitad dioses y mitad bestias, o mitad divinos y mitad diabólicos.

Este pensar en el «animal listo», inspirado en Darwin, era común a James y Dewey, pero James hizo más sitio que Dewey a la experiencia religiosa. James pensó que el tipo de unión entre el yo y el universo que describían los místicos era posible y deseable, pero no creía que se pudiera respaldar con argumentos filosóficos. Dejó de lado el Argumento desde el Holismo al Monismo y puso en su lugar un aprecio no argumentativo por los estados de conciencia especiales que habían descrito los místicos y otras personas religiosas. A Dewey, en cambio, no le interesaba la experiencia religiosa, salvo por una suerte de vago sentido panteísta, no-extático, de comunión con el universo (del tipo que describía Wordsworth).

James y Dewey escribieron antes del llamado giro lingüístico de la filosofía. Todavía hablaban de la experiencia más bien que del lenguaje. Los pragmatistas posteriores, en cambio, fueron influidos por Frege, Russell y Carnap. Figuras como Wilfrid Sellars y Donald Davidson aseveraron no solo que la única diferencia entre nosotros y los animales es nuestra capacidad de usar lenguaje, sino que, como lo expresó Sellars al anunciar su doctrina del «nominalismo psicológico»: todo percatarse de algo es un asunto lingüístico. Para Sellars, Davidson y yo mismo, los animales y los niños prelingüísticos pueden dar respuestas condicionadas a estímulos, pero no conocen nada. Tener conocimiento de algo es colocar enunciados sobre ese algo dentro de la práctica social de justificar, de dar y pedir razones.

Siguiendo esta línea de pensamiento, uno dirá, con Nelson Goodman, que no hay un Modo de Ser del Mundo. Repetirá también la tesis de James y Dewey de que la verdad no es una cuestión de correspondencia con la naturaleza intrínseca de la realidad, sino de dar con herramientas eficaces de auto-mejoramiento, donde este último es una cuestión de alcanzar más y mejores formas de felicidad. Desde este punto de vista, el problema del idealismo absoluto es que no ayuda a los proyectos de cooperación social (por ejemplo, construir una sociedad igualitaria, sin clases, sin castas), cuyo fomento debería ser la meta de la filosofía. Para los pobres y los que sufren no es de ninguna ayuda oír que la pluralidad es una ilusión.

Dewey pensaba que cierta idea que propusieron tanto Platón como Aristóteles había hecho muchísimo daño a Occidente: la de que hay una actividad llamada contemplación que es más verdaderamente humana —y más importante para una vida humana plena— que tales proyectos sociopolíticos. Pensaba que superar esa idea era esencial para revitalizar la filosofía occidental. (Puso como título *Reconstrucción en filosofía* a las conferencias que dio en Tokio inmediatamente después de la Primera Guerra Mundial. En mi opinión, esas conferencias conforman su mejor libro.) Aunque él mismo había sido en su juventud un idealista absoluto (periodo en el que también fue un ferviente cristiano), su giro hacia Darwin le llevó a considerar la búsqueda de autoconocimiento una desgraciada distracción de empresas que merecían más la pena.

Para los pragmatistas, como para Nietzsche, la idea toda de autoconocimiento —de averiguar qué es uno realmente, como quería Sócrates— es una mala idea. Hay que remplazar la ambición de autoconocimiento por la ambición de autocreación. El yo no es más que una red de creencias y deseos que se retejen a sí mismos continuamente. El sentido de pensar es agrandar esa red y hacerla más coherente, no hallar algo que hay por debajo y que la red presupone. Así que Nishida y Descartes y los griegos se equivocan al pensar que los seres humanos son criaturas específicamente intelectuales, donde «intelectual» se explica como «capaz de conocer cómo son realmente las cosas, de penetrar hasta la realidad

a través de la apariencia». Los humanos, como los demás animales, son buscadores de felicidad, y hablar del «intelecto humano» es simplemente un modo de referirse al hecho de que los humanos somos capaces de utilizar en esa búsqueda herramientas de las que carecen los animales sin lenguaje (y, por tanto, sin cooperación social).

Dewey habría estado de acuerdo con Kierkegaard en que la tradición intelectualista occidental que va desde los griegos hasta Hegel adopta como supuesto fundamental que «autoconocerse es conocer a Dios». Kierkegaard, como Nietzsche, protestó contra ese supuesto desde la perspectiva individualista que hemos dado en llamar existencialismo. Dewey, y escritores posteriores como Sellars y Davidson, lo hacen desde la perspectiva social del humanismo democrático.

Desde esta última, las grandes religiones del mundo han proporcionado consuelo a las víctimas de la injusticia social, pero no han hecho lo suficiente para cambiar la sociedad de modo que se minimice esa injusticia. La tarea de una filosofía completamente secularizada será contribuir a los esfuerzos por mejorar este mundo, en vez de posponer a otro la esperanza humana. Es útil pensar en Marx y Dewey como dos hombres que arrancaron como hegelianos, giraron a la Izquierda, y acabaron quedándose de Hegel solo lo que resultara compatible con Darwin.

Cuando deweyanos como yo leemos a Nishida o Nishitani, por supuesto que hay mucho con lo que podemos coincidir. En particular, podemos coincidir con su preferencia por Leibniz y Spinoza antes que por los Empiristas Británicos, y estamos de acuerdo en que estos últimos no comprendieron que la búsqueda de conocimiento es el intento de relacionar las cosas unas con otras. En cambio, recelamos profundamente de las «religiones de dependencia absoluta», expresión que Nishida utiliza para describir tanto el cristianismo como el budismo.[2]

El humanismo democrático que impulsaba a Dewey adopta como primer principio suyo que los seres humanos estamos solos: que no podemos buscar ayuda en nada salvo en nuestros semejantes. Tal

como lo expresó Yeats, «cuanto refulge en la noche lo alimenta / el resinoso corazón del hombre»[3]. Por eso, cuando Nishida dice que «las situaciones de extremo infortunio [...] están destinadas a despertar la conciencia religiosa en cada uno de nosotros», Dewey coincidiría con Nietzsche en que eso es una razón para recelar de la conciencia religiosa. Dewey insistiría en que los seres humanos deben juntarse para minimizar la posibilidad de situaciones de extremo infortunio, en vez de buscar consuelo en lo no-temporal.

Considérese el siguiente pasaje de la parte II de «The Logic of *Topos* and the Religous Worldview», de Nishida. Él escribe: «ciertamente no hay ninguna razón por la que los seres humanos tengan que ser religiosos, especialmente si uno postula la existencia social como el fundamento de la auto-existencia. Uno también podría lidiar con el problema de la vida y la muerte desde la perspectiva de la sociedad. Pero ¿no es la existencia humana lo que hay en la base de la sociedad?».[4] Interpreto que Nishida está diciendo que, aunque uno puede intentar remplazar la religión por la esperanza sociopolítica, el intento está condenado a fallar. No funcionará porque olvida la naturaleza del yo humano, la naturaleza de la existencia humana.

Dewey replicaría que no hay tal cosa como la naturaleza de la existencia humana o del yo humano. Los seres humanos son, aparte de sus atributos meramente biológicos, lo que la cultura hace de ellos. Puedes construir un nuevo tipo de ser humano creando un nuevo tipo de cultura (o, con Nietzsche, intentando con valentía convertirte en el Superhombre). Desde un punto de vista darwiniano, lo único que tienen los seres humanos que no es maleable por la cultura son esas neuronas extra que nos permiten usar el lenguaje.

Pero no hay nada filosóficamente interesante que aprender estudiando qué tienen en común todos los lenguajes humanos. Así que la filosofía no debe proponerse captar la naturaleza intrínseca de un objeto estable y continuo llamado la existencia humana, o el yo humano, o el lenguaje humano. Más bien, debería proponerse cambiar el lenguaje que usamos, desarrollando nuevas maneras de

hablar que nos permitan convertirnos en personas más felices. La transición de la Europa medieval a la moderna es un ejemplo de desarrollo de nuevas palabras que permitieron a la gente hacer cosas nuevas.

Dewey desde luego que desearía, en palabras de Nishida, lidiar con «el problema de la vida y la muerte desde la perspectiva de la sociedad». Tal como él veía la cuestión, ese problema había que resolverlo persuadiendo a los individuos de que vieran su vida como una contribución a un futuro humano utópico, y su muerte como redimida por la esperanza de que sus descendientes vivan en una sociedad mejor. Así como para Nishida, igual que para Kierkegaard, «solo quien conoce su propia muerte eterna sabe verdaderamente que es un individuo»,[5] para Dewey no hay muerte eterna si ha habido una vida que ha contribuido al bien social. En ese sentido, como en muchos otros, la existencia social es para Dewey «el fundamento de la auto-existencia». La esperanza social ocupa el lugar que Nishida reservaba a la Compasión.

Por decir todo esto de otro modo, Dewey negaría la tesis que Nishida formula en los siguientes términos: «en el fondo del yo hay algo que nos trasciende por completo, y ese algo, ni nos es extraño ni nos es externo».[6] Dewey replicaría que Nishida logra condensar limpiamente con su fórmula la doctrina central de la tradición metafísica occidental. Pero él quiere abandonar esa tradición. Hay que abandonarla, no porque haya sido refutada, sino porque el intento de construir un mundo mejor sobre la base de esa doctrina metafísica ha sido un experimento social fallido. Deberíamos darle la espalda sin más a ese experimento, no reapropiárnoslo (como Nishida y Heidegger querrían que hiciéramos).

Nishitani escribió que «allí donde hay seres finitos —y todas las cosas son finitas—, debe haber la nada; dondequiera que hay vida, debe haber muerte. Ante la faz de la muerte y de la nada, toda vida y experiencia pierden su certidumbre y su importancia como realidad, y a cambio pasan a parecer irreales. Desde tiempo inmemorial, el hombre ha estado continuamente expresando esta fugaz transitoriedad de la vida y la existencia, que la asemejan a un

sueño, a una sombra, o a la deslumbrante neblina del calor de verano».[7]

La reacción de Dewey a este pasaje sería: es verdad que esa comparación se ha hecho desde tiempo inmemorial, pero hacerla no ha producido ningún resultado útil. Ha sido un obstáculo al progreso social, más que una ayuda. Así que dejemos de hacer la comparación y démosle la vuelta al proceso. Pensemos mejor en lo infinito, lo completo, lo autocontenido, lo perfecto, como «una sombra, un sueño, una deslumbrante neblina». La actitud de Dewey es el epítome de lo que Nishitani llamó «ateísmo del progreso».

Yo no sé cómo resolver la disputa entre Nishitani y Dewey, y no lo intentaré en las conferencias que siguen. En lugar de ello, acentuaré sus discrepancias presentando una defensa del pragmatismo tan coherente como me sea posible. Una forma de acercar las tradiciones filosóficas de Oriente y Occidente es subrayar sus muchas similitudes. Yo haré lo contrario e intentaré hacer tan llamativas como pueda sus diferencias. Creo que la lucha en Occidente entre la metafísica, cuya culminación es el idealismo absoluto, y el ateísmo del progreso, cuya culminación es el pragmatismo lingüistificado de Sellars y Davidson, reproduce algunos aspectos de la tensión Oriente-Occidente. Confío en que esa tensión pueda clarificarse abogando una vez más por la causa de la tradición filosófica occidental que menos tiene en común con el pensamiento oriental.

---

[1] Kitarō Nishida, «The Logic of *Topos* and the Religious Worldview: Part II», trad. de Yusa Michiko, *The Eastern Buddhist* 20, n.º 1 (1987), p. 87.

[2] Kitarō Nishida, «The Logic of *Topos* and the Religious Worldview», trad. de Yusa Michiko, *The Eastern Buddhist* 19, n.º 2 (1986), p. 1.

[3] N. del T.: *Whatever flames upon the night / Man's own resinous heart has fed* («Two Songs from a Play»). Usamos la traducción de Antonio Rivero Taramillo, *Poesía reunida de W. B. Yeats*, Valencia, Pre-Textos, 2010.

[4] Kitarō Nishida, «The Logic of *Topos*», Part II, p. 82.

[5] Kitarō Nishida, «The Logic of *Topos*», p. 18.

[6] Kitarō Nishida, «The Logic of *Topos*», Part II, p. 85.

[7] Keiji Nishitani, *Religion and Nothingness*, trad. de Jan Van Bragt. Berkeley: University of California Press, 1982, p. 7.

Diseño de cubierta:
Anaí Miguel

Título original:
On Philosophy and Philosophers. Unpublished Papers, 1960-2000

© Esta edición ha sido publicada en lengua inglesa por Cambridge University Press en 2020

Edición en formato digital: 2024

© de la edición, W. P. MAŁECKI y CHRIS VOPARIL, 2024
© de la traducción, ÁNGEL MANUEL FAERNA GARCÍA-BERMEJO, 2024
© EDITORIAL TECNOS (GRUPO ANAYA, S.A.), 2024
Calle Valentín Beato, 21
28037 Madrid
www.tecnos.es

ISBN ebook: 978-84-309-8997-3

Está prohibida la reproducción total o parcial de este libro electrónico, su transmisión, su descarga, su descompilación, su tratamiento informático, su almacenamiento o introducción en cualquier sistema de repositorio y recuperación, en cualquier forma o por cualquier medio, ya sea electrónico, mecánico, conocido o por inventar, sin el permiso expreso escrito de los titulares del Copyright.